

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 22, 2005, 18:00:21 ; Search time 824 Seconds

(without alignments)  
4580.514 Million cell updates/sec

Title: US-10-757-093-4

Perfect score: 3354

Sequence: 1 MKFLGLSLSLAAPSISLGRP.....RKPKAAASHLRARMTSIDKN 634

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 5544816 seqs, 2976611598 residues

Total number of hits satisfying chosen parameters: 11089632

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool.p/US10757093/unat\_18032005\_164457\_27782/app\_query.fasta\_1.775  
-DB=Published Applications NA -QFMT=fastcap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPEL=0 -LOOPEXT=0 -INITS=bites -START=1 -END=1 -MATRIX=blosum62  
-TRANS=humand0.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=200 -THR MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODELOCAL -OUTFMT=ptco -NORM=ext -HEARSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10757093\_QCGN\_1\_1723@unat\_18032005\_164457\_27782  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FEAPOP=6 -FEAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
20: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
21: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
22: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1717.5	51.2	3451	15	US-10-161-403-108
2	1717.5	51.2	3451	18	US-10-161-408-20
3	1717.5	51.2	14627	15	US-10-161-403-109
4	1717.5	51.2	14627	18	US-10-161-408-21
5	1715	51.1	8654	18	US-10-640-422-160
6	1714.5	51.1	8654	10	US-09-845-064-11
7	1714.5	51.1	8654	10	US-09-845-064-11
8	1714.5	51.1	9390	10	US-09-845-064-22
9	1714.5	51.1	9390	10	US-09-845-064-18
10	1714.5	51.1	9390	10	US-09-845-064-20
11	1714.5	51.1	9688	10	US-09-845-064-50
12	1714.5	51.1	15208	10	US-09-845-064-51
13	1714	51.1	10011	18	US-10-680-824A-19
14	1714	51.1	12224	16	US-10-336-566-83
15	1714	51.1	32798	16	US-10-424-638-1
16	1712	51.0	9285	10	US-09-845-064-52
17	1712	51.0	15077	10	US-09-845-064-57
18	1711	51.0	1809	18	US-10-432-777-16
19	1711	51.0	1812	15	US-10-161-403-105
20	1711	51.0	1812	17	US-10-365-493-24529
21	1711	51.0	1812	18	US-10-161-408-16
22	1711	51.0	11096	19	US-10-872-156-9
23	1709	51.0	1812	18	US-10-149-533A-13
24	1709	51.0	4652	10	US-09-893-525-36
25	1709	51.0	4652	18	US-10-763-380-36
26	1706	50.9	5390	10	US-09-893-525-41
27	1706	50.9	5390	18	US-10-763-380-41
28	1706	50.9	5418	10	US-09-893-525-38
29	1706	50.9	5418	18	US-10-763-380-38
30	1703	50.8	11978	10	US-09-792-568-8
31	1703	50.8	12438	10	US-09-792-568-9
32	1680.5	50.2	2755	17	US-10-239-907A-43
33	1680.5	50.1	4341	17	US-10-239-907A-47
34	1673.5	49.9	2141	17	US-10-663-241-27
35	1673.5	49.9	6975	17	US-10-431-252-17
36	1672.5	49.9	2001	17	US-10-239-907A-38
37	1672.5	49.9	2001	18	US-10-332-406A-21
38	1672.5	49.9	2001	18	US-10-800-161-29
39	1672.5	49.9	2730	17	US-10-239-907A-41
40	1672.5	49.9	4072	17	US-10-239-907A-49
41	1672.5	49.9	12817	18	US-10-332-406A-22
42	1672.5	49.9	13274	18	US-10-332-406A-20
43	1671.5	49.8	4129	16	US-10-322-656-28
44	1671	49.8	4280	15	US-10-027-880-2
45	1671	49.8	4309	15	US-10-027-880-4

#### ALIGNMENTS

RESULT 1  
US-10-161-403-108  
; Sequence 108, Application US/10161403  
; Publication No. US20030119104A1  
; GENERAL INFORMATION:  
; APPLICANT: Perkins, Edward  
; APPLICANT: Perez, Carl  
; APPLICANT: Lindenbaum, Michael  
; APPLICANT: Greene, Amy  
; APPLICANT: Leung, Josephine  
; APPLICANT: Fleming, Elena  
; APPLICANT: Stewart, Sandra  
; APPLICANT: Shellard, Joan  
; TITLE OF INVENTION: CHROMOSOME-BASED PLATFORMS  
; FILE REFERENCE: 24601-420  
; CURRENT APPLICATION NUMBER: US/10/161,403  
; CURRENT FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: 60/294,758  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: 60/366,891

```

PRIOR FILING DATE: 2002-03-21
NUMBER OF SEQ ID NOS: 129
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 108
LENGTH: 3451
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: HindIII Fragment containing the beta-glucuronidase
OTHER INFORMATION: coding sequence, the rDNA intergenic spacer, and
OTHER INFORMATION: the MastI sequence
US-10-161-403-108

Alignment Scores:
Pred. No.: 1.15e-169 Length: 3451
Score: 1717.50 Matches: 338
Percent Similarity: 68.84% Conservative: 95
Best Local Similarity: 53.74% Mismatches: 175
Query Match: 51.21% Indels: 21
DB: Gaps: 11

US-10-757-093-4 (1-634) x US-10-161-403-108 (1-3451)
QY 13 AlaAlaProSerLeuGlyThrProAlaAlaArgHisPheProArgAsnGlu----- 29
DB 1221 TCTGCACCGGATCTCGAGATC-----GAATTCCTCCGCGCGGGAATTCCTAG 1268
QY 30 --MetThrGlnHisGlnProLeuValLeuValArgProGlnArgThrSerArg 48
DB 1269 TGGATCCCGCGGTACGGTCCCTTATG--TTACGTCCTGTAGAAACCCCAACCCGT 1325
QY 49 GluLeuValAsnLeuAspGlyLeuThrPheAlaLeu--AlaSerGlyLeuAsnAsp 67
DB 1326 GAATCAAAAATTCGACGGCTGTGGCATTCAGTGGATGCGGAAACCTGTGAATT 1385
QY 68 ThrAlaGlnProThrAlaProLeuProGlyLeuGluCysProValProAlaSer 87
DB 1386 GACACCGCTTGGCGGAACGGGTTACAGAAAGCCGCGCAATGCTGTGCCAGGACGT 1445
QY 88 TyrAsnAspIlePheIleSerArgGluIleHisAspHisValGlyTyrValTyrGln 107
DB 1446 TTTAACGATCAGTTCGCGCATGACGATATTCGTAAATATGTGGCAACGCTGTATCA 1505
QY 108 ArgGluValIleValProGlySerGlyTyrPheGlnGluArgTyrLeuValArgAlaGlnSer 127
DB 1506 CGGAGGCTTTATACGAAAGGTGGGCGGCCAGGCTATCGCTGCGCTTCGATGCG 1565
QY 128 AlaThrHisHisGlyArgGlyLeuValAsnAsnArgLeuValAlaGlnHisValGly 147
DB 1566 GTCACTCATTAACGCAAAAGTGGGTCAATATCAGGAAGTGAAGCATCAGGCGGCG 1625
QY 148 TyrThrProPheGluValAspValThrGlyLeuValAlaProGlyGlyCysPheArgLeu 167
DB 1626 TATATACGCAATTTGAAGCGCATGTCAACCCGTATGTTATTCGCGGAAAGTGAACGTATTC 1685
QY 168 ThrIleGlyValAsnAsnGluLeuThrHisGlyThrIleProProGlyLysIleThrThr 187
DB 1686 ACGATTGTGTGAACAGAACTGAATGCGACAGCTATCCCGCGGAAAGTGAATACC 1745
QY 188 GluAsnAlaThrGlyLysArgGlyLeuThrTyrGlnHisAspPheTyrAsnTyrAlaGly 207
DB 1746 ---GACAAAGAACGCAAGAAAGACATCTTCACTTCATGATTTCTTAATCAAGCGCGG 1802
QY 208 LeuAlaArgSerIleTyrPheValProGlnGlnHisIleGlnAspIleThrVal 227
DB 1803 ATCCATGCGACGCGTAATGCTTCAACACCGCAACCTGCGTGAAGCATATCACCGTG 1862
QY 228 ValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGlyValAla 245
DB 1863 GTGACGATGTCGCGCAAGCTGTACCAACCGCTGTGACGCGAGTG---GTGCGC 1919
QY 246 AsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAla 265

```

```

DB 1920 AAT-----GGTATGTCAGCGTTGAACTGCGTGAATGGCGATCAACAGGTGTTGCA 1970
QY 266 LysAlaSerGlyAlaGlnGlyThrValThrIleProSerValLysLeuTyrGlnProGly 285
DB 1971 ACTGACAAAGGACACGCGGAGCTTGTGAAATCCGCACCTTGGCAACCGGCT 2030
QY 286 AlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValAlaAsp 305
DB 2031 GAAGTTATCTCTAATGAACGTAGCTC-----ACAGCAAAAGCCAGACAGATGTGAT 2084
QY 306 ThrTyrAsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIle 325
DB 2085 ATCTACCCGCTGCGCGTCCGTCAGTGTGAGTGAAGGGGGAACAGTTCCTGATC 2144
QY 326 AsnGlyLysProPheThrPheThrGlyPheGlyLysHisGlyAspThrAlaValArgGly 345
DB 2145 AACCAACAAACGCTTCTACTTTCGTGGCTTGGCGTCAATGAAGATCGGATTTGGCGGCG 2204
QY 346 LysGlyHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTyrIleGlyAla 365
DB 2205 AAGGATTCGATAAGCTGCTGATGTGTCACATCAACGATTATGACTGATGGGCGC 2264
QY 366 AsnSerPheArgThrSerHisTyrProTyrAlaGlnGluValMetAspPheAlaAspArg 385
DB 2265 AACTCTTACCTGATCCTCGCATTAACCTTACGTTGAAGATGCTGACTGGCGACATGAA 2324
QY 386 AsnGlyIleValValIleAspGlnThrProAlaValAlaGlyLeuAsnIleAlaLeu--Met 404
DB 2325 CATTGATGCTGTGATGATGATTAACCTGACGCTGTGCGCTTAACTCTCTTATGACAT 2384
QY 405 GlyValSerGlnSerGlyAlaPro--GlnThrPheThrProAspAlaIleAsnAspLys 423
DB 2385 GGTTCGAAGCGGGAACAAAGCGAAAGATCTACAGCAAGAGAGGAGTCAACCGGGA 2444
QY 424 ThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysHisAla 443
DB 2445 ACTCAGACGCCACCTTACAGCGATTAAGCGATGAAGCGTATGCGCGTGAACAAACCCCA 2504
QY 444 SerValIleMetTyrPheSerIleAlaAsnGluProAlaSerHisGlyAspGlyAlaArgGlu 463
DB 2505 AGCGTGTATGTGATGATGATTTGCCAAGCAACCGATACCCGTCGCCAAGTGCACGCGAA 2564
QY 464 TyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPhe 483
DB 2565 TATTTGCCCACTGGCGGAAACAAAGCGTAATCTGAATCCGACGCGTCCGATTCACCTGC 2624
QY 484 AlaSerValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspAlaSer 503
DB 2625 GTCMAATGATATGTTCTGCACGCTCAACCCGATACCATGCGATCTCTTATATGTCTG 2684
QY 504 CysIleAsnArgTyrPheGlyTyrPheSerGlnThrGlyAspLeuGlnGluAla 523
DB 2685 TGCCTAACCCTTATATACGTTGATGTCCAAAGCGGAGATTTGGAAACGCAAGAAAG 2744
QY 524 AlaLeuGluLysGlnLeuHisGlyTyrGlnGlyLysPheHisArgProIleValMetThr 543
DB 2745 GTACTGAAAGAAAGAACTTGCCTGCGTCGACAGAAACTGCATCAGCGGATTCATCAACC 2804
QY 544 GluTyrGlyAlaAspThrLeuAlaGlyLysHisSerIleLeuGlyLeuProTyrPheSerGlu 563
DB 2805 GAATACGGGCGGTATGCTTACCGCGGTCGCTCAATATATACACGCACTGTGAGTGA 2864
QY 564 GluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAspArgIleGlnSerMet 583
DB 2865 GAGTATCAGTGTGATGCTGATATGATACACCGCGTCTTGTATGGCTGACGCGCGTC 2924
QY 584 AlaGlyGlnHisValTyrAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgVal 603
DB 2925 GTGCGTGAACAGGTATGAAATTTGCGCAATTTGCAACCTCGCAAGCGCATATGTCGCCGT 2984
QY 604 AspGlyAsnLysLysGlyValPheThrArgAspArgLysProLysAlaIleHisSer 623
DB 2985 GCGCGTATCAAGAAAGGAGATTTTACCCGCGACCGCAACCGAAAGTCGGGCGCTTTTCTG 3044

```

QY 624 LeuArgAlaArgThrPheSerIleLeu 632  
 DB 3045 CTGCAGAAAAGCTGGACCTGGCATGAAAC 3071

RESULT 2  
 US-10-161-408-20  
 ; Sequence 20, Application US/10161408  
 ; Publication No. US20040214290A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Perez, Carl  
 ; APPLICANT: Fabijanski, Steven  
 ; APPLICANT: Perkins, Edward  
 ; TITLE OF INVENTION: Plant Artificial Chromosomes  
 ; FILE REFERENCE: 24601-419  
 ; CURRENT APPLICATION NUMBER: US/10161,408  
 ; PRIOR FILING DATE: 2002-05-30  
 ; PRIOR APPLICATION NUMBER: US 60/294,687  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: US 60/236,329  
 ; NUMBER OF SEQ ID NOS: 51  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 20  
 ; LENGTH: 3451  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: HindIII Fragment containing the beta-glucuronidase  
 ; OTHER INFORMATION: coding sequence, the rDNA intergenic spacer, and  
 ; OTHER INFORMATION: the Mael1 sequence  
 US-10-161-408-20

Alignment Scores:  
 Pred. No.: 1.15e-169 Length: 3451  
 Score: 1717.50 Matches: 338  
 Percent Similarity: 68.84% Conservative: 95  
 Best Local Similarity: 53.74% Mismatches: 175  
 Query Match: 51.21% Indels: 21  
 Gaps: 11

US-10-757-093-4 (1-634) x US-10-161-408-20 (1-3451)

QY 13 AlaAlaProSerLeuGlyThrProAlaAlaArgHisPheProArgLeu----- 29  
 DB 1221 TCTGCACCGGATCTCGAGATC-----GAATTCCTCCGCGCGCGCAATTCTAG 1268

QY 30 ---MetThrGlnHisGlnProLeuIleValArgProGlnArgThrSerArg 48  
 DB 1269 TGGATCCCGGCGTACGCTCCCTTATG---TTACGCTCTGTAGAAACCCCAACCGT 1325

QY 49 GluLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu---AlaSerGlyLeuAsnAsp 67  
 DB 1326 GAAATCAAAAACCTGACGCGCTGTGGCATTGAGTTCGATCGCAAACTGTGAAAT 1385

QY 68 ThrAlaGlnProTrpThrAlaProLeuProLysIleuGluCysProValProAlaSer 87  
 DB 1386 GAGCAGCGCTGTGGGAAACGCGCTTACAGAAACCGCGCAATTGCTGTGCGAGGAGT 1445

QY 88 TyrAsnAspIlePheIleSerArgGluIleHisAspHisValGlyTrpValTyrTrpGln 107  
 DB 1446 TTAAACGATACGATTCGCCCATGCAAGATATTCGTAATATGTGGCAACGCTGTGATCAG 1505

QY 108 ArgGluValIleValProLysGlyTrpSerGlnGluArgTrpLeuValArgAlaGluSer 127  
 DB 1506 CGGAGATCTTTATACGAAAGGTGGCGACGAGCGTANCGCTGCGTTTCGATGCG 1565

QY 128 AlaThrHisHisIleGlyArgIleTyrValAsnAsnArgLeuValAlaGluHisValGly 147  
 DB 1566 GTCAATCATTAACGCAAAAGTGGGTCAATATATCAGAAATGATGAGCATCAGGGCGC 1625

QY 148 TyrThrProheGluAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeu 167

DB 1626 TATACCGCATTTGAACCCATGTCACGCCGTATGTTATTCGCCGAAAGTGTACGATC 1685  
 QY 168 ThrIleGlyValAsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThr 187  
 DB 1686 ACAGTTGTGTGAACAACGAACTGAACCTGACAGACTATCCCGCGGAATGTGATTAAC 1745

QY 188 GlyAsnAlaThrGlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGly 207  
 DB 1746 ---GACGAAACCGCAAGAAAGAGCTCTTACTTCATATATTTCTTAATCTACGCCGG 1802

QY 208 LeuAlaArgSerIleTrpLeuTyrSerValProGlnGlnHisIleGlnAspIleThrVal 227  
 DB 1803 ATCCATCGCAGCGTAATAGCTCTACACAGCGCAACACCTGGGTGACATATACCGGTG 1862

QY 228 ValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGluValAla 245  
 DB 1863 GTACACGATGTCGCGGACGCTGAACACGCGCTGTGACCTGCGCAGGTG---GTGGCC 1919

QY 246 AsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAla 265  
 DB 1920 AAT-----GATGATGACCGTTGAATCCTGATCCGATCAACAGGTGTTGCA 1970

QY 266 LysAlaSerGlyAlaGlnGlyThrValThrIleProSerValLysLeuTrpGlnProGly 285  
 DB 1971 ACTGACAAAGGACACAGCGGACCTTGCAAGTGATGATCCGACCTTCGCAACCGGCT 2030

QY 286 AlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValAlaAsp 305  
 DB 2031 GAAGGTATCTCTATACAACTGACGTC-----ACAGCCAAAGCCAGACAGATGTAT 2084

QY 306 ThrTyrAsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIle 325  
 DB 2085 ATCTACCGCTGCGCGCTCGGATCCGATCGGATGAGGAGGCGCAACGTTCTGATC 2144

QY 326 AsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGluAspThrAlaValArgGly 345  
 DB 2145 AACCAACAAACCGTTACTTACTTACGCTTGGCGCGTATGAATGCGATTGCGCGCGC 2204

QY 346 LysGlyHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTrpIleGlyAla 365  
 DB 2205 AAAGGATTCGATACGCTGATGATGTCAGATGACGATCAACGATTAATGACCTGATGGGCGC 2264

QY 366 AsnSerPheArgThrSerHisTyrProTyrAlaGluGluValMetAspPheAlaAspArg 385  
 DB 2265 AACTCTACCGTACCTGCGCTTACCTTACGCTGAAGATGCTCGACTGGGAGATGAA 2324

QY 386 AsnGlyIleValValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu---Met 404  
 DB 2325 CATGCGATCGGTGATGATGAACTGACAGCTGCTGCGCTTTAACCTCTTTAGCAT 2384

QY 405 GlyValSerGluSerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLys 423  
 DB 2385 GGTTCGAAACCGGCAACAAAGCTGAAGAACTGTACAGCAAGGACGATCAACCGGGAA 2444

QY 424 ThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAla 443  
 DB 2445 ACTCAGAGCGGACCTTACAGCGGATTAAGACTGATTAACGCGGTGCAAAAACCAACCA 2504

QY 444 SerValValMetTrpSerIleAlaAsnGluProAlaSerHisIleGluAspGlyAlaArgGlu 463  
 DB 2505 AGCGTGATATGTCAGATATGTCAGCAAGAACCGGATACCGCTCCGACAGTGCACGGAA 2564

QY 464 TyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPhe 483  
 DB 2565 TATTTGCGCGCAGCTGCGGACGACGAGCAACGTAACCTGATCCGACGCTCGACACCTGC 2624

QY 484 AlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSer 503  
 DB 2625 GTCAATGTAATGTTTCGCAAGCTCAACCGATACATTCAGCATCTTTGATAGTCTG 2684

QY 504 CysIleAsnArgTyrPheGlyTrpTyrSerGlnThrGlyAspLeuGluAlaGluAla 523

Db 2685 TGCCGTAACCGTATATACGGTGTGATGTCCAAAGCGGCGCATTTGGAAACGCGACAGAG 2744  
 Qy 524 AlaleuGluYsGluLeuHhIGLYTTPGInguYsPheHIsarGProIleValMetThr 543  
 Db 2745 GTACTGGAAAAAGAACTTCTGGCTGGGAGAGAAATGCAATCGCATATATATATATACC 2804  
 Qy 544 GluTYrGluYAlaAspThrLeuAlaGlyLeuHIsSerIleLeuGlyLeuProTTPSerGlu 563  
 Db 2805 GAATACGCGCTGGATACGTTAGCGGGCTGCACCTCAATGTACACCGACATGTGAAGTAA 2864  
 Qy 564 GluPheGluValAlaMetLeuAspMetTYrHIsarGValPheAspArgIleGluSerMet 583  
 Db 2865 GAGATTCAGTGTGATGAGCTGATATGTATATACCGCGCTTGTATCCGTCAGCGGCTC 2924  
 Qy 584 AlaGlyGluHIsValITTPAspPheAlaAspPheGluHIsValIleLeuArgVal 603  
 Db 2925 GTCCGTGAACAGATATGAAATTTGCCGATTTTGGACCTTGCAGAGGCAATATCCGCTT 2984  
 Qy 604 AspGlyAsnIlySlyGlyValPheThrArgAspArgIlyProIlyAlaAlaHIsSer 623  
 Db 2985 GCGCGTAACAGAAAGGGGATCTTCACCGCGACCGCAACGAAAGTGGCGGCTTTCTG 3044  
 Qy 624 LeuArgAlaArgTTPHrSerIleAsp 632  
 Db 3045 CTGCAGAAACCGCTGACTGGCATGAAC 3071

RESULT 3  
 US-10-161-403-109  
 ; Sequence 109, Application US/10161403  
 ; Publication No. US20030119104A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Perkins, Edward  
 ; APPLICANT: Perez, Carl  
 ; APPLICANT: Lindenbaum, Michael  
 ; APPLICANT: Greene, Amy  
 ; APPLICANT: Leung, Josephine  
 ; APPLICANT: Fleming, Elena  
 ; APPLICANT: Stewart, Sandra  
 ; APPLICANT: Sheilard, Joan  
 ; TITLE OF INVENTION: CHROMOSOME-BASED PLATFORMS  
 ; FILE REFERENCE: 24601-420  
 ; CURRENT APPLICATION NUMBER: US/10/161,403  
 ; PRIOR FILING DATE: 2002-05-30  
 ; PRIOR APPLICATION NUMBER: 60/294,758  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: 60/366,891  
 ; NUMBER OF SEQ ID NOS: 129  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 109  
 ; LENGTH: 14627  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: pag11a Plasmid  
 US-10-161-403-109

Alignment Scores:  
 Pred. No.: 9.01e-169 Length: 14627  
 Score: 1717.50 Matches: 338  
 Percent Similarity: 68.84% Conservative: 95  
 Best Local Similarity: 53.74% Mismatches: 175  
 Query Match: 51.21% Indels: 21  
 DB: 15 Gaps: 11

US-10-757-093-4 (1-634) x US-10-161-403-109 (1-14627)  
 Qy 13 AlaAlaProSerLeuGlyThrProAlaAlaArgHIsPheProArgAsnGlu----- 29  
 Db 12069 TCTGCACCGGATCTCAGATC-----GAATTCGCGGCGCGGAATTCATCAG 12116  
 Qy 30 ---MetThrGlnHIsGluInProLeuIleYsValArgProGluArgThrSerArg 48  
 Db 30 ---MetThrGlnHIsGluInProLeuIleYsValArgProGluArgThrSerArg 48

Db 12117 TGGATCCCGCGGTACGGTCACTCCCTTATG---TTACGTCCTGTAGAAACCCCAACCGGT 12173  
 Qy 49 GluLeuValAsnLeuAspGlyLeuTTPYsPheAlaLeu---AlaSerGlyLeuAsnAsp 67  
 Db 12174 GAATCAAAAACCTCAGACCGGCTGTGGGCAATTCAGCTGTGATCGCAAAATCTGTGAATT 12233  
 Qy 68 ThrAlaGluProThrHrAlaProLeuProIlySlyLeuGluCysProValProAlaSer 87  
 Db 12234 GACACGCTGTGTGGGAAAGCGGTTACAAAGAAAGCGGGAATTCCTGTGTGAGGCACT 12293  
 Qy 88 TYrAsnAspIlePheIleSerArgGluIleHIsPheHIsValGlyTYrValTYrGln 107  
 Db 12294 TTTAACGATCAGTTCCCGCATGACATATTCGAATTATGTGGGCAACCTCGTATCAG 12353  
 Qy 108 ArgGluValIleValProIlySlyTYrPserGlnGluArgTYrLeuValArgAlaGluSer 127  
 Db 12354 CGGAAAGCTTTATATACGAAAGGTGGGAGGCGACACCGCTATCGTGTGGTTCGATGCG 12413  
 Qy 128 AlaThrHIsHIsGlyArgIleTYrValAsnAspArgLeuValAlaGluHIsValGlyGly 147  
 Db 12414 GTACATCTATACGCAAAAGTGGGTCATATACAGAAATGATGAGCATCAGGCGGC 12473  
 Qy 148 TYrThrProPheGluAlaAspValThrGluLeuValAlaProGlyGluYsPheArgLeu 167  
 Db 12474 TATACCCCATTTGAAGCCGATGTACAGCCGATGTATGTATGCGGGAAGTATCATCTATC 12533  
 Qy 168 ThrIleGlyValAsnAsnGluLeuThrHIsGluTYrIleProProGlyYsAlaIleThr 187  
 Db 12534 ACAATTGTGTGAACAACAGTAACTGCGACACTATCCCGCGGAATGTGTATCC 12593  
 Qy 188 GluAsnAlaThrGlyYsArgIleGlnTYrGlnHIsAspPheTYrAsnTYrAlaGly 207  
 Db 12594 ---GACGAAACCGCAAGAAAGAGCTTACTTCATGATTTTTHAATCAACCCGG 12650  
 Qy 208 LeuAlaArgSerIleTTPLeuTYrSerValProGlnGlnHIsIleGluAspIleThrVal 227  
 Db 12651 ATCCATCGAGGATGATGCTTACACACCGCAACCTGGGTGAGATATCACCGGT 12710  
 Qy 228 ValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTYrGluValGluValAla 245  
 Db 12711 GTGACGATGTGGCGCAAGACTGTAACACAGCGGTGTGTGACTGGCAGAGTG---GTGCC 12767  
 Qy 246 AsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaAlaValAla 265  
 Db 12768 AAT-----GGATGTCAAGCGTGAACCTGATGCGATGCGATCAACAGGTGTTC 12818  
 Qy 266 LysAlaSerGlyAlaGlnGlyThrValThrIleProSerValYsLeuTTPGlnProGly 285  
 Db 12819 ACTGACAAAGGACCAAGCGGACTTTGCAAGTGTGAATCCGCACTTGGCAACCGGGT 12878  
 Qy 286 AlaAlaTYrLeuTYrGlnLeuGlnValAsnIleValIleGlySerSerGlyAspValValAsp 305  
 Db 12879 GAAGTTATCTCTTATGAACGTGACTC-----ACAGCAAAAGCCACAGAGTGTAT 12932  
 Qy 306 ThrTYrAsnLeuAlaThrGlyValArgThrValYsValAlaGlySerGlnPheLeuIle 325  
 Db 12933 ATCTACCGCTGCGGCGGTGCGGATCGGTCAAGTGCAGAGGAGGCAACAGTCTCTGATC 12992  
 Qy 326 AsnGlyYsProPheTYrPheThrGlyPheGlyYsHIsGluAspThrAlaValArgGly 345  
 Db 12993 AACCAAAACCGTTCTTACTGAGCTGTGGCTGTGAAGATGGCGATTTGGCGGC 13052  
 Qy 346 LysGlyHIsAspProAlaTYrMetValHIsAspPheGlnLeuMetIlySTPPIleGlyAla 365  
 Db 13053 AAAGGATTCGATTAACGTGCTGATGTGACAGATCAACGATTTAAAGCATGTGAGGCGC 13112  
 Qy 366 AsnSerPheArgThrSerHIsTYrProTYrAlaGluGluValMetAspPheAlaAspArg 385  
 Db 13113 AACTCTAACCTGATCTGCAATTAACCTTAACGCTAAGAGATGCTCGAGTGGCGAGATGAA 13172  
 Qy 386 AsnGlyIleValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu---Met 404  
 Db 13173 CATGCACTGTGTGATGTATGAATGCACTGTGCGCTTTAACTCTTTTAAAGGCAATT 13232



```
QY 405 GYValSerGluSerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLys 423
DB 13233 GGTTCGAAAGCCGGCAACAAGCCGMAAAAGCTGACAGCCAAAGGCGACGCGGAA 13292
QY 424 ThrGlnGlnAlaIleValGlnAlaIleArgGluLeuIleAlaArgAspLysAsnIleAla 443
DB 13293 ACTCAGACAGCGGCACTTACAGGCGATTAAAGACTGATACCGGTGACAAAACCCCA 13352
QY 444 SerValIleMetTrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGlu 463
DB 13353 AGCGTGATGATGCGAGTATTCGCAACGAAACCGGATCCCGCCGCAAGGTGACCGGAA 13412
QY 464 TyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPhe 483
DB 13413 TATTTCGCGCCGACGCGGCAAGCAAGCGTAACTGATCCGACGCGCTCGATCACCCTGC 13472
QY 484 AlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSer 503
DB 13473 GTCAATGTAATGTTCTGCGACGCTCACACCGATACATCAGCATCTTGTGATGTCGTG 13532
QY 504 CysIleAsnArgTyrPheGlyTyrTyrSerGlnThrGlyAspLeuGluAlaGluAla 523
DB 13533 TGCCTAACCGTTATTAACGTTGATGTCGCAAGCGCGCATTTGGAAACGCGACAGAG 13592
QY 524 AlaLeuGlnLysGluLeuHisGlyTyrGlnGlnLysPheHisArgProIleValMetThr 543
DB 13593 GATCTGAAATAAACACTTCGCGCTGCGAGGAAATGATGATACCGCATTCATCATCAC 13652
QY 544 GluTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyrPheSerGlu 563
DB 13653 GAATACGGCGGTGATACGTTAGCGCGGCTGCATCATGTAACCCAGCATGTGAGAGTGA 13712
QY 564 GluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAspArgIleGluSerMet 583
DB 13713 GAGTATCAGTGTCATGCTGCGATGATGATACCCGCGCTTGTATGCGCGTCCGACG 13772
QY 584 AlaGlnGlnIleValIleThrAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgVal 603
DB 13773 GTCCGGTAAACAGGATGAAATTCGCGGATTTTCGCGCTGCGAGGCAATATTCGCGGT 13832
QY 604 AspGlyAsnLysLeuGlyValPheThrArgAspArgGlyAspProLysAlaAlaHisSer 623
DB 13833 GCGCGTAAACAAGAGGAGATCTTCAACCGCGACCGCAACCGAAGTCGCGGCTTTTCTG 13892
QY 624 LeuArgAlaArgTyrThrSerIleAsp 632
DB 13893 CTGCAAAAACGCTGACCTGCGCATGAAC 13919

RESULT 4
US-10-161-408-21
; Sequence 21, Application US/10161408
; Publication No. US20040214290A1
; GENERAL INFORMATION:
; APPLICANT: Perez, Carl
; APPLICANT: Pabjanek, Steven
; APPLICANT: Perkins, Edward
; TITLE OF INVENTION: Plant Artificial Chromosomes, Uses thereof, and Methods of Prepara
; FILE REFERENCE: 24601-419
; CURRENT APPLICATION NUMBER: US/10/161,408
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: US 60/294,687
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: US 60/296,329
; PRIOR FILING DATE: 2001-06-04
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PasteSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 14627
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
```

```
OTHER INFORMATION: pag1a Plasmid
US-10-161-408-21

Alignment Scores:
Pred. No.: 9,01e-169 Length: 14627
Score: 1717.50 Matches: 338
Percent Similarity: 68.84% Conservative: 95
Best Local Similarity: 53.74% Mismatches: 175
Query Match: 51.21% Indels: 21
DB: 18 Gaps: 11

US-10-757-093-4 (1-634) x US-10-161-408-21 (1-14627)

QY 13 AlaAlaProSerLeuGlyThrProAlaAlaArgHisPheProArgAsnGlu----- 29
DB 12069 TCTGACCCGGATCTCCAGATC-----GAAATCCCGCGCGCGCAATTCCTAG 12116
QY 30 ---MetThrGlnHisGluGlnProLeuIleValArgProGlnArgThrSerArg 48
DB 12117 TGGATCCCGCGGATACGCTGATCCCTTATG---TTACGTCTGTGAAACCCCAACCGT 12173
QY 49 GluLeuValAsnLeuAspGlyLeuTyrPheAlaLeu---AlaSerGlyLeuAsnAsp 67
DB 12174 GAATCAAAAACCTCGACGCGCTGTGCGCATTCAGTCTGATGCGCAAACTGTGGAATT 12233
QY 68 ThrAlaGlnProThrAlaProLeuProLysGlyLeuGluCysProValSer 87
DB 12234 GACACGCTGGTGGGAAAGCGGCTTAACAAGAGCCGCGCAATGCTGTGCGACGCGAT 12293
QY 88 TyrAsnAspIlePheIleSerArgGluIleHisAspHisValGlyTyrValTyrTyrGln 107
DB 12294 TTTAACGATAGTATCCCGCATGCGATGATTCGTAATTATGTGGCAACGTCTGTATCAG 12353
QY 108 ArgGlnValIleValProLysGlyTyrPheSerGlnArgTyrGlyLeuValArgAlaGluSer 127
DB 12354 CCGGAAGTCTTTTATACCGAAAGGTGGGACGCGACGATGCTGCTGCTTCGATGCG 12413
QY 128 AlaThrHisHisGlyArgIleTyrValAsnAsnArgLeuValAlaGlnHisValGlyGly 147
DB 12414 GTCACTATTTACCGCAAGAGTGGGTCAATATACAGAAATGATGAGATCAGGCGGCG 12473
QY 148 TyrThrProPheGlnAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeu 167
DB 12474 TATACCGCATTTGAACCGCATGTACAGCGCTATGTTATGCGGAAAGATGATGATC 12533
QY 168 ThrIleGlyValAsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThr 187
DB 12534 ACAGTTTGTTGTAACAACGAACGAACTGACGACATATCCCGCGGAAATGATGATTAAC 12593
QY 188 GlyAsnAlaThrGlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGly 207
DB 12594 ---GACGAAACCGGCAAGAAAGAGTCTTATTCATGATATTTCTTAACTACGCGGG 12650
QY 208 LeuAlaArgSerIleTrpLeuTyrSerValProGlnGlnHisIleGlnAspIleThrVal 227
DB 12651 ATCCATCGACGCGTATATGCTCTACACACCGCAACACCTGGGTGACGATATCACCGTG 12710
QY 228 ValIleAspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGluValAla 245
DB 12711 GTGACCCAGTCCCGCAAGCTGTACACACGCGCTGTGTAATGCGGACGAGTG---GTGCGCC 12767
QY 246 AsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAla 265
DB 12768 AAT-----GATATGCAAGCTGGAATGCGTGAATGCGGATCAACAGTGGTGGCA 12818
QY 266 IysAlaSerGlyValaGlnGlyThrValThrIleProSerValLysLeuTrpGlnProGly 285
DB 12819 ACTGACACGACCAACGCGGACTTTGCAAGTGTGAATCCGCACTTGTGCAACCGGGT 12878
QY 286 AlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValaGlySerSerGlyAspValValAsp 305
DB 12879 GAAGTTATCTATGAACTGTACGTC-----ACAGCCAAAAGCCAGACAGAGTGAT 12932
```

```

QY 306 ThrTyrAsnLeuAlaThrGlyValArgThrValLeuValAlaGlySerGlnPheLeuIle 325
DB 12933 ATCTACCCGGCTGGCCGCGGATCCGCTCACTGCACTGAAAGGCGAAGCTTCTATC 12992
QY 326 AsnGlyLysProPheTyrPheThrGlyPheGlyLysIleGluAspThrAlaValArgGly 345
DB 12993 AACCAACAAACCGTTCTACTTACTGCTTGGCCGTCATGAAGATGCGGATTTGCGCGGC 13052
QY 346 LysGlyIleAspProAlaTyrMetValHisAspPheGlnMetLeuIleSTPIIleGlyAla 365
DB 13053 AAGAGATTCCAGTAAACGCTGATGTCACGATCACCATTAAGACTGATGATGGGCC 13112
QY 366 AsnSerPheArgThrSerHisTyrProTyrAlaGluGluValMetAspPheAlaAspArg 385
DB 13113 AACCTCCACCGTACTGCTGATACCTTACGCTGAAGAAGATGCTCCGACTGGCAGATGAA 13172
QY 386 AsnGlyIleValValIleAspGluThrProAlaValIleLeuAsnIleAlaLeu--Met 404
DB 13173 CATGGCATCGTGGATGATGAAGAACTGACAGCTGCTTAACTCTTAAAGCATTTAGGCAT 13232
QY 405 GlyValSerGluSerGlyAlaPro--GlnThrPheThrProAspAlaIleAsnAspLys 423
DB 13233 GATTTCACACGGCGCAACACCGCAAGAACTGACAGCAAGAGCATCAACGGGGA 13292
QY 424 ThrGlnGluAlaHisIleGlyGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAla 443
DB 13293 ACTCAGACGGCGCACTTAACAGGGCATTAAGAGCTGATTAACGGCTGACAAAACCAACCA 13352
QY 444 SerValValMetTyrSerIleIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGlu 463
DB 13353 AGCGTGTGATGATGGAGATATTTGCCAACGAAACCGATCCCGTCCGCAAGGTGACCGGAA 13412
QY 464 TyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPhe 483
DB 13413 TATTCGCGGCATCTGGCGGAAGCAACCGTAACCTGATCCGACGCTCCGATCATCTGC 13472
QY 484 AlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSer 503
DB 13473 GTCAGTATGATGTTCTGCGACGCTCACACCGATACCATGCGATCTTTGATGTGCTG 13532
QY 504 CysIleAsnArgTyrPheGlyTyrTyrSerGlnThrGlyAspLeuGluGluAlaGluAla 523
DB 13533 TGCCTGAACCGTTATTAAGGTGTGATGTCAAAGCGCGCATTTGAAACGCGAGAGAG 13592
QY 524 AlaLeuGluLysGluLeuHisGlyTyrGlnGluLysPheHisArgProIleValMetThr 543
DB 13593 GTACGAAAAAAGAACTTCTGGCGTGGCAGAGAAACTGCATACGCCGATATCATCAC 13652
QY 544 GlyTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyrPheSer 563
DB 13653 GAATACGGCGGTGATAGTTAGCGGCGTGCATCAATGTACACCGCATGTGAGTGA 13712
QY 564 GluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAspArgIleGluSerMet 583
DB 13713 GAGTATCAGTGTGCATGCGTGAATGTATCACCGGCTCTTATATGCGGTACACGGCGTC 13772
QY 584 AlaGlyLysIleValIleAspPheAlaAspPheGlnThrAsnLeuGlyIleIleArgVal 603
DB 13773 GTGCGTGAACAGTATGAAATTTCCCGATTTTGCACCTCGCAAGGCATATTTCCCGTT 13832
QY 604 AspGlyAsnLysGlyValPheThrArgAspArgLysProLysAlaAlaHisSer 623
DB 13833 GCGCGTAAACAAAGGAGATCTTACCCGCGAACCGAAAGTCGCGGCTTTTCTG 13892
QY 624 LeuArgAlaArgThrPheSerIleAsp 632
DB 13893 CTGCAAAACGCTGACTGCGCATGAAC 13919

RESULT 5
US-10-640-422-160
; Sequence 160, Application US/10640422
; Publication No. US20040229229A1
; GENERAL INFORMATION:

APPLICANT: Cheo, David
APPLICANT: Brasch, Michael A.
APPLICANT: Temple, Gary F.
APPLICANT: Hartley, James L.
APPLICANT: Byrd, Devon R.N.
TITLE OF INVENTION: Use of Multiple Recombination Sites with Unique Specificity in
FILE REFERENCE: 0942.5010004
CURRENT APPLICATION NUMBER: US/10/640,422
CURRENT FILING DATE: 2003-08-14
PRIOR APPLICATION NUMBER: US 60/402,920
PRIOR FILING DATE: 2002-08-14
PRIOR APPLICATION NUMBER: US 60/169,983
PRIOR FILING DATE: 1999-12-10
PRIOR APPLICATION NUMBER: US 60/188,020
PRIOR FILING DATE: 2000-03-09
PRIOR APPLICATION NUMBER: US 09/732,914
PRIOR FILING DATE: 2000-12-11
NUMBER OF SEQ ID NOS: 183
SOFTWARE: PatentIn version 3.0
SEQ ID NO 160
LENGTH: 5898
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: pms/GW
US-10-640-422-160

Alignment Scores:
Pred. No.: 4,53e-169 Length: 5898
Score: 1715.00 Matches: 338
Percent Similarity: 68.68 Conservative: 94
Best Local Similarity: 53.74 Mismatches: 173
Query Match: 51.13 Indels: 24
DB: 18 Gaps: 11

US-10-757-093-4 (1-634) x US-10-640-422-160 (1-5898)
QY 12 LeuAlaAlaProSerLeuGlyThrProAlaAlaArg--HisPheProArgAsnGluMet 30
DB 1263 TTAAACCATCCCAAGTTTGTAACAAAAGAGGCTCATTTAACTTTAAGAGAGATATAT 1322
QY 31 ThrGlnHisGluGlnProLeuIleLysValArgProGlnArgThrSerSerArgGluLeu 50
DB 1323 ACCATG-----GTCGCTCTCTGTGAAGAACCCCAACCGGTGAATTC 1361
QY 51 ValAsnLeuAspGlyLeuThrLysPheAlaLeu-----AlaSerGlyLeuAsnAsp 67
DB 1362 AAAAACTGGACGGCTGTGGCATTCAGTTCGGAATCGCAAACTGTGAATTGATCAG 1421
QY 68 ThrAlaGlnProTyrThrAlaProLeuProLysGlyLeuGluLysProValProAlaSer 87
DB 1422 -----CGTTGTGGGAAAGCGGCTTACAGAAAGCGGCAATTGCTGTGCCAGGCACT 1475
QY 88 TyrAsnAspIlePheIleSerArgGluIleHisAspHisValGlyTyrValTyrTyrGln 107
DB 1476 TTAAACGATCACTTGGCGCATGACAGATATTCGATTTATGCGGGAAGCTGTGATCAG 1535
QY 108 ArgGluValIleValProLysGlyTyrPheGlnGluArgTyrLeuValArgAlaGluSer 127
DB 1536 CCGGAAGTCTTATTCGGAAGGTTGGCAGGCGCAGCGATCTGCTCGTTTCGATCAG 1595
QY 128 AlaThrHisLeuGlyAlaArgIleTyrValAsnAspArgLeuValAlaGluHisValGly 147
DB 1596 GTACCTCATTTACGGCAAGGTGGGTCAATATTCAGAAAGTATGAGCATCAGGCGGC 1655
QY 148 TyrThrProPheGluAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeu 167
DB 1656 TATACGCCATTTGAAGCCGATGTCAAGCGGATGTATTTGCGGGAAGAGTACGATATC 1715
QY 168 ThrIleGlyValAsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThr 187
DB 1716 ACCGTTTGTTGAACAAAGAACTGAATCTGCACTATCCCGCCGGAATGCTGATTTACC 1775

```



```

FEATURE:
NAME/KEY: promoter
LOCATION: (5557)..(5771)
OTHER INFORMATION: No. US20030175976A1alaline synthetase promoter
FEATURE:
NAME/KEY: misc_feature
LOCATION: (5771)..(5818)
OTHER INFORMATION: MCS multiple cloning site
FEATURE:
NAME/KEY: gene
LOCATION: (5818)..(7717)
OTHER INFORMATION: GUS gene coding for beta glucuronidase
FEATURE:
NAME/KEY: polyA_signal
LOCATION: (7718)..(8447)
OTHER INFORMATION: Poly A from 35S ribosome
FEATURE:
NAME/KEY: misc_feature
LOCATION: (8447)..(8474)
OTHER INFORMATION: MCS multiple cloning site
FEATURE:
NAME/KEY: misc_feature
LOCATION: (8474)..(8647)
OTHER INFORMATION: T-DNA right border
US-09-845-064-11

```

## Alignment Scores:

Pred. No.:	8.83e-169	Length:	8654
Score:	1714.50	Matches:	337
Percent Similarity:	69.12%	Conservative:	95
Best Local Similarity:	53.92%	Mismatches:	177
Query Match:	51.12%	Indels:	17
DB:	10	Gaps:	9

US-10-757-093-4 (1-634) x US-09-845-064-11 (1-8654)

```

QY 15 ProSerLeuGlyThrProAlaAlaArgHisPheProArgAsnGluMetThrGlnHisGlu 34
DB 5742 CCCAAGCTTGGCCGGCGCTTAACACGCGTGAATCCTT--AATTAAGTCGACTAGGGGCT 5799
QY 35 GlnProLeuIleuValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAsp 54
DB 5800 GGTACGTCCTTATGTTAGCTGCTGTAAGAAACCCCAACCCGTAATCAAAAAAATCGAC 5859
QY 55 GlyLeuTrpLysPheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnPro 71
DB 5860 GGCCTGGGCAATTCAGTCTGATCGCGAAACGCTGGAAATGATGACG-----CGTTGG 5913
QY 72 TrpThrAlaProLeuProLysGlyLeuGluCyProValProAlaSerTyraAsnAspIle 91
DB 5914 TGGGAAAGCCCGCTTAACAAGAACCCGGCAATTGCTGTGCCAGCAGTTTAAACGATCAG 5973
QY 92 PheIleSerArgGluIleHisAspHisValGlyTrpValTyrrGlnArgGluValIle 111
DB 5974 TTGGCCGATCAATATTCGTAATATGCGGGCAACGCTGTAATCAAGCGCAAGCTTT 6033
QY 112 ValProLysGlyTrpSerGlnGluArgTyrrLeuValArgAlaGluSerAlaThrHis 131
DB 6034 ATACCGAAAGTTGGGACGCGACGCGATGCTGTGCTGCTTCGATGCGGTCACCTCATTTAC 6093
QY 132 GlyArgIleTyrrValAsnAsnArgLeuValAlaGlnHisValGlyGlyTyrrTrpProPhe 151
DB 6094 GGGAAAGTGGGTCAATATCAAGAAAGTATGAGATGAGAGCGCGCTAATACGCAATTT 6153
QY 152 GlnAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyVal 171
DB 6154 GAAGCCGATTCACGCGGTATGTTATGCGCGGAAAAAGTATACGATACCCGTTGTGTG 6213
QY 172 AsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThr 191
DB 6214 AACCAACGAACTGAAGTCGACAGACTATCCGCCGGAATGATGATTC--GACGAAAC 6270
QY 192 GlyLysArgIleGlnThrTyrrGlnHisAspPheTyraAsnTyrrAlaGlyLeuAlaArgSer 211

```

```

DB 6271 GGCAGAAAAAGACGCTTAATCTTCATGATTTCTTAACATACCGGAATCCATCCAGC 6330
QY 212 IleTrpLeuTyrrSerValProGlnGlnHisIleGlnAspIleThrValAlaThrAspVal 231
DB 6331 GTAATGCTCTACACACCGGAAACACCTGGGTGACGATATCACCTGTGTGACGATGTC 6390
QY 232 -----AspGlyAspAsnGlyLeuIleAsnTyrrGluValGlyValAlaAsnGlnThrThr 249
DB 6391 GCCAGAGACTGTATACCAACCGGTCTGTGACTGGCAGAGT--CTGCCAAT----- 6438
QY 250 GlyGlnIleGlnIleSerValIleAspGluAspGlyValIleValAlaLysAlaSerGly 269
DB 6439 GGTGATGTACGCGTTGAATGCGGTGATGCGGATCAACAGGTGGTTCAACTGGACAGGC 6498
QY 270 AlaGlnGlyThrValThrIleProSerValLysLeuThrGlnProGlyAlaAlaTyrrLeu 289
DB 6499 ACTAGCGGACCTTTGAGAGTGGAATCCGCACTTGGCAACCGGTAAGGTTATCTC 6558
QY 290 TyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValValAspThrTyraAsnLeu 309
DB 6559 TATGAATGTGGCTC-----ACAGCCAAAGCCACAGAGTGATATCTACCCGCTT 6612
QY 310 AlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysPro 329
DB 6613 CGCGTCGCGATCCGTCAGTGGCAGTGAAGGGGCAACAGTCTCTGATTAACCAACAAACCG 6672
QY 330 PheTyrrPheThrGlyPheGlyLysHisGluAspThrAlaValArgGlyLysGlnHisAsp 349
DB 6673 TTCTACTTTACCTGCGCTTGTGTGTGTCATGAAAGTCGCAAGCTTGGTGGCAAGATTCGAT 6732
QY 350 ProAlaTyrrMetValHisAspPheGlnLeuMetLysTyrrIleGlyValaAsnSerPheArg 369
DB 6733 AACGTCTAGTGTGTCACACCAACCGCATTAATGAGATGTGATTTGGGCAACTCTACCT 6792
QY 370 ThrSerHisTyrrProTyrrAlaGluGluValMetAspPheAlaAspArgAsnGlyIleVal 389
DB 6793 ACCTCCATTAACCTTACCTTAACCTGAAGATGCTCGATGGGCAAGTAAATGACATGCAACG 6852
QY 390 ValIleAspGlyThrProAlaValGlyLeuAsnIleAlaLeu--MetGlyValSerGlu 408
DB 6853 GTGATGATGAACCTGCTGCTGCTGCTTAACCTCTCTTAAGCATTTGGATTTCCGAAGG 6912
QY 409 SerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAla 427
DB 6913 GGCACAAACCGAAAGAACTGTACAGCAAGAGGCGTCAACGGGGAACCTCAGCAAGCG 6972
QY 428 HisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMet 447
DB 6973 CACTTACAGGCGATTAAGAGCTGATVAGCGGTGACCAAAACCAACCAACGCGTGTGATG 7032
QY 448 TrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGluTyrrPheGluPro 467
DB 7033 TGGAGATTTGCCACAGCAACCGATACCCGTCGCAAGGTGCAACGGGAATATTTCCGGCCA 7092
QY 468 LeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGly 487
DB 7093 CTGGCGGAGAACCGGTAACCTCGACCCGACGCGTCGATACACCTGCTCATATGTAATG 7152
QY 488 ThrAlaThrTyrrGlnLeuAspArgIleSerAspLeuPheAspValSerValIleAsnArg 507
DB 7153 TTCTGCGAGCTCACACCGATACCATACAGCATCTTTGAATGTGTGCTGTAACCT 7212
QY 508 TyrPheGlyTyrrTyrrSerGlnThrGlyAspLeuGluGluAlaGluAlaIleLeuGluLys 527
DB 7213 TATTAAGATGTGATATCCAAAGCGGATTTGGAAACGCGCAAGAGATCTGAAAAA 7272
QY 528 GlnLeuHisGlyTyrrGlnGluLysPheHisArgProIleValMetThrGluTyrrGlyAla 547
DB 7273 GAACCTTCGCTGGGAGGAGAACTGATCAGCGCATATATCATCATCAACCAATACGCGCGTG 7332
QY 548 AspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyrrSerGluGluPheGlnVal 567

```

Db 7333 GATACGTTAGCCGGCGTCACTCATGTAGACCCGACATGTGAGTGAAGATCACTAGTGT 7392  
Qy 568 GIMetLeuapbMetYrhiAargValPhasaparglIegIuserMetAlaGlyGluHis 587  
Db 7393 GCATGGCTGATGATGATACCGCGGTCTTGATCGCTACGCCGCTCGTCGGTGAACAG 7452  
Qy 588 ValTrrasbPheAlaAepPheGlnThrAsnLeuGlyIleIleAargValAaspGlyAsnLys 607  
Db 7453 GTATGGAATTTCCGCGATTTTTCGACCTCGCAAGCATATTTGCCGCTTGGCGGTAAACAG 7512  
Qy 608 LygGlyValbheThrAargAargLysProLysAlaAlaHisSerLeuAargAlaArg 627  
Db 7513 AAAGGGATCTTCACTCCGCGAACCGCAAGTCGGCGCTTTCTGTGCAAAAAGCGC 7572  
Qy 628 TrpThrSerLeuAep 632  
Db 7573 TGGACTGGCATGAAC 7587

RESULT 7  
US-09-845-064-14  
; Sequence 14, Application US/09845064  
; Publication No. US20030175976A1  
; GENERAL INFORMATION:  
; APPLICANT: MERISTEM THERAPEUTICS  
; TITLE OF INVENTION: CLEAN SYNTHETIC VECTORS, PLASMIDS, TRANSGENIC PLANTS  
; TITLE OF INVENTION: CLEAN SYNTHETIC VECTORS, PLANTS CONTAINING SAID VECTORS, AND THEIR  
; TITLE OF INVENTION: METHODS OF PRODUCTION  
; FILE REFERENCE: SYNVEC1  
; CURRENT APPLICATION NUMBER: US/09/845,064  
; CURRENT FILING DATE: 2001-04-27  
; NUMBER OF SEQ ID NOS: 57  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 14  
; LENGTH: 8654  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Plasmid  
; OTHER INFORMATION: pmr1196  
; FEATURE:  
; NAME/KEY: rep origin  
; LOCATION: (1) .. (654)  
; OTHER INFORMATION: Origin of replication ori RK2  
; FEATURE:  
; NAME/KEY: rep origin  
; LOCATION: (655) .. (1263)  
; OTHER INFORMATION: Origin of replication ori COLEI  
; FEATURE:  
; NAME/KEY: gene  
; LOCATION: (1264) .. (2603)  
; OTHER INFORMATION: NPT III gene coding for neomycin  
; OTHER INFORMATION: phosphotransferase and Kanamycin resistance  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (2604) .. (4098)  
; OTHER INFORMATION: TrfA locus from RK2 coding for two proteins, P285  
; OTHER INFORMATION: and P382, enabling the increase of the rate of  
; OTHER INFORMATION: replication  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4105) .. (4271)  
; OTHER INFORMATION: T-DNA left border  
; FEATURE:  
; NAME/KEY: terminator  
; LOCATION: (4272) .. (4559)  
; OTHER INFORMATION: No. US20030175976A1:line synthetase terminator  
; FEATURE:  
; NAME/KEY: gene  
; LOCATION: (4560) .. (5556)  
; OTHER INFORMATION: NPT II gene coding for neomycin phosphotransferase  
; OTHER INFORMATION: and Kanamycin resistance  
; FEATURE:  
; NAME/KEY: promoter

LOCATION: (5557) .. (5771)  
; OTHER INFORMATION: No. US20030175976A1:line synthetase promoter  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (5771) .. (5818)  
; OTHER INFORMATION: MCS multiple cloning site  
; FEATURE:  
; NAME/KEY: gene  
; LOCATION: (5818) .. (7717)  
; OTHER INFORMATION: GUS gene coding for beta glucuronidase  
; FEATURE:  
; NAME/KEY: polyA signal  
; LOCATION: (7718) .. (8447)  
; OTHER INFORMATION: Poly A from 35S ribosome  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (8447) .. (8474)  
; OTHER INFORMATION: MCS multiple cloning site  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (8474) .. (8647)  
; OTHER INFORMATION: T-DNA right border  
US-09-845-064-14

Alignment Scores:  
Pred. No.: 8,836-169 Length: 8654  
Score: 1714.50 Matches: 337  
Percent Similarity: 69.12% Conservative: 95  
Best Local Similarity: 53.92% Mismatches: 177  
Query Match: 51.12% Indels: 17  
DB: 10 Gaps: 9

US-10-757-093-4 (1-634) x US-09-845-064-14 (1-8654)

Qy 15 ProSerLeuGlyThrProAlaAlaArgHisPheProAargLysGluMetThrGlnHisGlu 34  
Db 5742 CCGAAGCTTGGCGCGCGCTTAAACACCGGTGATCTT--AATTATCGACTAGAGGCT 5799  
Qy 35 GlnProLeuIleLysValAargProGlnAargThrSerSerArgGluLeuAasp 54  
Db 5800 GGTCACTCCCTTGTGTTACGCTCTGTAGAAACCCCAACCGGTAAATCAAAAATCTGCAC 5859  
Qy 55 GlyLeuTrpLysPheAlaLeu-----AlaSerGlyLeuAsnAepThrAlaGlnPro 71  
Db 5860 GGCCTGTGGGCAATTCAGTCTGATCCGGAATACTGGATTCATG-----CGTTGG 5913  
Qy 72 TrpThrAlaProLeuProLysGlyLeuGluCysProValProAlaSerTyraAsnAasp 91  
Db 5914 TGGGAAAGCGCTTAAACAAAGACCGGCAATTGCTGTCCAGGAGTTTAAACATCAG 5973  
Qy 92 PheIleSerArgGluIleHisAaspHisValGlyTrpValTyrrGlnAargGluVal 111  
Db 5974 TTGCGCGATGCAATATTCGTAATTTATGCGGCAACGCTGATATCAGCGCAAGCTTT 6033  
Qy 112 ValProLysGlyTrpSerGlnGluAargTyrrLeuValAargLysSerAlaThrHis 131  
Db 6034 ATACCGAAGGTTGGGAGCGCCAGCGTATGCTGTGCTGCTTCATGCGGCTCATCTTAC 6093  
Qy 132 GlyArgLysValAsnAaspValAlaGluHisValGlyGlyTyrrThrProPhe 151  
Db 6094 GGCAAGATGGGATCATATATCAAGAAAGTATGAGATCAGGCGGCTATACGCCATTT 6153  
Qy 152 GluAlaAaspValThrGluLeuValAlaProGlyGlyLysPheAargLeuThrIleGlyVal 171  
Db 6154 GAAGCCGATGTCAAGCGATATTTATTTGCGGGAAGAAAGTATGATACACCGTTTGTG 6213  
Qy 172 AsnAsnGluLeuThrHisGluThrIleProGlyLysIleThrThrGlyAsnAlaThr 191  
Db 6214 AACACGAACTGAACGTGACGACTATCCCGCGGAATGATGATTAC--GACGAAAC 6270  
Qy 192 GlyLysArgLysGlnThrGlnHisAaspPheTyraAsnTyrrAlaGlyLeuAlaArgSer 211  
Db 6271 GGCAGAAAGAAAGCAGCTTATCTTCATGATTTCTTAACTATCCGGAATCATTCGACG 6330



OTHER INFORMATION: No. US20030175976a1aline synthetase promoter  
FEATURE:  
NAME/KEY: promoter  
LOCATION: (5369)..(6111)  
OTHER INFORMATION: Enhanced promoter from 35S ribosome  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (6111)..(6159)  
OTHER INFORMATION: MCS multiple cloning site  
FEATURE:  
NAME/KEY: gene  
LOCATION: (6159)..(8050).  
OTHER INFORMATION: GUS gene coding for beta glucuronidase  
FEATURE:  
NAME/KEY: polyA\_signal  
LOCATION: (8051)..(8780)  
OTHER INFORMATION: Poly A from 35S ribosome  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (8780)..(8807)  
OTHER INFORMATION: MCS multiple cloning site  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (8807)..(8980)  
OTHER INFORMATION: T-DNA right border  
US-09-845-064-22

Alignment Scores:  
Pred. No.: 9.32e-169 Length: 8987  
Score: 1714.50 Matches: 337  
Percent Similarity: 69.12% Conservative: 95  
Best Local Similarity: 53.92% Mismatches: 177  
Query Match: 51.12% Indels: 17  
DB: 10 Gaps: 9

US-10-757-093-4 (1-634) x US-09-845-064-22 (1-8987)

QY 15 ProSerLeuGluThrProAlaAlaArgHisPheProArgHisGluMetThrGlnHisGlu 34  
DB 6075 CCGAAGCTTGGCCGGCGGTTAAACACGCGTGATCTT- -AATTAAAGTCACTTAGGGGT 6132  
QY 35 GlnProLeuLeuValArgProGlnArgThSerSerArgGluLeuValAsnLeuAsp 54  
DB 6133 GGTACGTCCTTATGTTACGTCCTGTTGAAACCCCAACCGGTGAACAAAACCTGAC 6192  
QY 55 GlyLeuTrpLysPheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnPro 71  
DB 6193 GGCCTGGGCGATTCAGTTCGATCGGAAACCTGCGAATTGATCAG-----CGTTGG 6246  
QY 72 TrpThrAlaProLeuProLysGlyLeuGluCysProValProAlaSerTyraAsnAspIle 91  
DB 6247 TGGGAAAGCCGCTTACAAAGAACCGGGCAATTGCTGTGCCAGCGACTTTTAACGATCAG 6306  
QY 92 PheIleSerArgGluLeuHisAspHisValGlyTrpValTyrrGlnArgGluValIle 111  
DB 6307 TTGGCCGATGCAATATTCGTAATATGCGGGCAACGCTCGTATTCAGCCGGAAGTCTTT 6366  
QY 112 ValProLysGlyTrpSerGlnArgTyrrLeuValAlaGluHisValGlyGlyTyrrThrProPhe 151  
DB 6367 ATACCGAAAGGTGGCGACGCGCGATGCTGCTGCTTCGATCGGTCACCTCATAC 6426  
QY 132 GlyArgLeuTyrrValAsnAsnArgLeuValAlaGluHisValGlyGlyTyrrThrProPhe 151  
DB 6427 GGCAGAGTGGGCAATTAATCAAGAAAGATGAGATCGAGCGGCGCTATACGCCATTT 6486  
QY 152 GluAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyVal 171  
DB 6487 GAAGCCATGTCAACCGCTATGTTATGCGGGAAAGTAGACTATACCCGTTTGCTG 6546  
QY 172 AsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThr 191  
DB 6547 AACACGAACTGAACCTGCGACACTATCCCGCGGGAATGGATTAACC--GACGAAAC 6603

QY 192 GlyLeuArgGlnIleThrTyrrGlnHisAspPheTyraAsnTyrrAlaGluLeuAlaArgSer 211  
DB 6604 GCGAAGAAAAGAGCTTATCTTCATGATTTCTTTAAGTATCCGGAATCCATCCGAC 6663  
QY 212 IleTrpLeuTyrrSerValProGlnGlnHisIleGlnAspIleThrValValThrAspVal 231  
DB 6664 GTAAGCTCTACACACGCGGAACACCTGGGTGAGATATCAACCGGTGAGCATGTC 6723  
QY 232 -----AspGlyAsnAsnGlyLeuIleAsnTyrrGluValGluValAlaAsnGlnThr 249  
DB 6724 GCGCAAGACTGTAAACACGCGCTGTGACTGGCAGGTG--GTGGCAAT----- 6771  
QY 250 GlyGlnIleGlnIleSerValIleAspGluAspGlyValAlaIleValAlaLysAspGly 269  
DB 6772 GGTGATGTCAGCGTGAATCTGCTGATCGGATCAACAGGTGTTGCAATGAGCAAGGC 6831  
QY 270 AlaGlnGlyThrValThrIleProSerValLysLeuTrpGlnProGlyAlaAlaTyrrLeu 289  
DB 6832 ACTAGCGGACTTTGCAAGTGTGAATCCGCACTCTGGCAACCGGGGTGAAGGTTATCTC 6891  
QY 290 TyrrGlnLeuGlnValAsnIleValGlySerSerGlyAspValValAspThrTyrrAsnLeu 309  
DB 6892 TATGAACCTGTGCGTC-----ACAGCCAAAGCCAGACAGATGTATATCTACCCGCT 6945  
QY 310 AlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysPro 329  
DB 6946 CCGGTGGCATCCGCTGACGTGAGCGAGGAGGCGGAACAGTCTTGATTAACCAAAACCG 7005  
QY 330 PheTyrrPheThrGlyPheGlyLysHisGluAspThrAlaValArgGlyLysGlnHisAsp 349  
DB 7006 TTCTACTTACTGGCTTGCTGCTGATGAAGATCGGCACTTGGCGCAACGATTCGAT 7065  
QY 350 ProAlaTyrrMetValHisAspPheGlnLeuMetLysTrpIleGlyValAsnAspPheArg 369  
DB 7066 AACGTGCTGATGTGACGACGACGCACTTAATGATGATGATGTGGCGCAACTCTACCGT 7125  
QY 370 ThrSerIleTyrrProTyrrAlaGluGluValMetAspPheAlaAspArgAsnGlyIleVal 389  
DB 7126 ACCTCGATTAACCTTACGCTGTAAGATGCTGATCGGCAATGACATGACATGCGATCGT 7185  
QY 390 ValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu--MetGlyValSerGlu 408  
DB 7186 GTATGATGATGAACGTGCTGCTGCTTAACTCTCTTAGACATGTGTTGCAAGCG 7245  
QY 409 SerGlyAlaPro--GlnThrPheThrProAspAlaIleAsnAspTyrrThrGlnAla 427  
DB 7246 GGCACAAAGCCGAAAGAACTGACAGCGAAGAGCGACTCAACGGAACTCGACAGCG 7305  
QY 428 HisLeuGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAlaSerValIleMet 447  
DB 7306 CACTTACAGCGATTTAAAGCTGATAGCGCGTGAACAAACCAACCCAGCGTGTGATG 7365  
QY 448 TrpSerIleAlaAsnGluProAlaSerHisGluAspGlyValaArgGluTyrrPheGluPro 467  
DB 7366 TGGAGTATTCGCAACGAAACCGGATACCCGTCGCAAGGTGACAGGGAATTTTCGCCCA 7425  
QY 468 LeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGly 487  
DB 7426 CTGCGGAAACCAACGGGTAAATCGAACCCGACGCTCGGATCACTCGCTCAATGTATG 7485  
QY 488 ThrAlaThrTyrrGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArg 507  
DB 7486 TTCTGCAAGCTCAACACCGATACATACAGGATCTTTGATGTGCTGTGCTCGAACCGT 7545  
QY 508 TyrrPheGlyTyrrTyrrSerGlnThrGlyAspLeuGluGluAlaGluAlaLeuGluLys 527  
DB 7546 TATTAACGATGATGTCGCAACGCGGATTTGGAACCGCAGAGAAAGTACTGGAAGAAA 7605  
QY 528 GlyLeuHisGlyTyrrGlnGluLysPheHisArgProIleValMetThrGlyTyrrGlyAla 547  
DB 7606 GAACCTTGTGCGCTGGCAGGAGAACTGCATCAACCGAATTAATCAACCGAATTCGGGCTG 7665  
QY 548 AspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProThrPheArgGluPheGlnVal 567



[illegible]



```
OTHER INFORMATION: and kanamycin resistance
FEATURE:
NAME/KEY: promoter
LOCATION: (5560)..(5771)
OTHER INFORMATION: No. US20030175976A1:line synthetase promoter
FEATURE:
NAME/KEY: promoter
LOCATION: (5772)..(6514)
OTHER INFORMATION: Enhanced promoter from 35S ribosome
NAME/KEY:
LOCATION: (6514)..(6554)
OTHER INFORMATION: MCS multiple cloning site
FEATURE:
NAME/KEY: gene
LOCATION: (6554)..(8453)
OTHER INFORMATION: GUS gene coding for beta glucuronidase
FEATURE:
NAME/KEY: polyA_signal
LOCATION: (8454)..(9183)
OTHER INFORMATION: Poly A from 35S ribosome
FEATURE:
NAME/KEY: misc feature
LOCATION: (9183)..(9210)
OTHER INFORMATION: MCS multiple cloning site
FEATURE:
NAME/KEY: misc feature
LOCATION: (9210)..(9383)
OTHER INFORMATION: T-DNA right border
US-09-845-064-20
```

```
Alignment Scores:
Pred. No.: 9.92e-169 Length: 9390
Score: 1714.50 Matches: 337
Percent Similarity: 69.12% Conservative: 95
Best Local Similarity: 53.92% Mismatches: 177
Query Match: 51.12% Indels: 17
DB: Gaps: 9
```

US-10-757-093-4 (1-634) x US-09-845-064-20 (1-9390)

```
QY 15 ProSerLeuGlyThrProAlaAlaArgHisPheProArgAsnGluMetThrGlnHisGlu 34
Db 6478 CCCAAGCTTGGCCGGCCGCTTAACACGCGTGATCCTT-AAATTAAGTCGACCTAGGGGT 6535
QY 35 GlnProLeuIleLysValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAsp 54
Db 6536 GGTCAAGTCCCTTATGTACGTCCTGTAGAAACCCCAACCCGCTGAATAAATAAATCGAC 6595
QY 55 GlnLeuTrpLysPheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnPro 71
Db 6596 GGGCTGTGGGCAATTCAGTCTGATCGGAAAACTGTGAATTGATCG-----CGTTGG 6649
QY 72 TrpThrAlaProLeuProLysGlyLeuGluCysProValProAlaSerTyraAsnAspIle 91
Db 6650 TGGGAAAGCCGCTTAACAAGAACCGGCAATTCCTGTCAGGCACTTTTAACGACAG 6709
QY 92 PheIleSerArgGluIleHisAspHisValGlyTrpValTyryGlnArgGluValIle 111
Db 6710 TTGGCCGATGCAATATTCGAATTATGCGGGCAACGCTGATCGATCGCGCAAGTCTTT 6769
QY 112 ValProLysGlyTrpSerGlnGluArgTyryLeuValArgAlaGluSerAlaThrHisHis 131
Db 6770 ATACCGAAGAGGTGGGAGCGGACGCTATCGTCTGCTGCTTCGATCGCGCTACCTATAC 6829
QY 132 GlyArgIleTyryValAsnAspArgLeuValAlaGluHisValGlyGlyTyryThrProPhe 151
Db 6830 GGCAGAAAGTGGCTCATATACAGAAAGTGAATGAGATCAGGCGGCTATACGCCATTT 6889
QY 152 GlnAlaAspValThrGluLeuValAlaProGlyLysPheArgLeuThrIleGlyVal 171
Db 6890 GAAGCGATGTCAAGCCGATGTATTGTCGCGGAAAGTGAATGATACACCGTTTGTGTG 6949
```

```
QY 172 AsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThr 191
Db 6950 AACACGAACTGAACCTGACAGACTATCCCGCGGAAAGTGTATACC---GACGAAAC 7006
QY 192 GlyLysArgIleGlnThrTyryGlnHisAspPheTyraSerTyryAlaGlyLeuAlaArgSer 211
Db 7007 GGCAGAAAGAAAGCACTTATTCATTCATGATTTCTTAATCAATGCGGATCATCGACAGC 7066
QY 212 IleTrpLeuTyrySerValProGlnGlnHisIleGlnAspIleThrValValThrAspVal 231
Db 7067 GTAAATGCTTACACACCGGCAACCTGGGTGAGCATATCACCGTGTGACGATGTC 7126
QY 232 -----AspGlyAspAsnGlyLeuIleAsnTyryGluValAlaGluAlaAsnGlnThrThr 249
Db 7127 GCGCAAGACTGTAAACACGCGCTGTGATCGGACGTC---GTGGCCAT----- 7174
QY 250 GlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAlaLysAlaSerGly 269
Db 7175 GGTGATGTCAAGCGTTGAACCTGCGTATGCGGATCAACAGTGTGTGCAACTGACAGAGC 7234
QY 270 AlaGlnGlyThrValThrIleProSerValLysLeuTrpGlnProGlyValAlaLysIleu 289
Db 7235 ACTAAGGGGACCTTTCAGATGTAATCGCACCTCTGGCAACCGGTGAGCGTTATCTC 7294
QY 290 TyryGlnLeuGlnValAsnIleValGlySerSerGlyAspValValAspThrTyryAsnLeu 309
Db 7295 TATGAACGTGGCTC-----ACAGCCAAAGCCAGACAGTGTGATATCACCGGCTT 7348
QY 310 AlaThrGlyValAlaArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysPro 329
Db 7349 CCGGTGCGGATCCGGTCAGTGCAGTGAAGGCGGAAACGTTCTCATTAACCAACAAACCG 7408
QY 330 PheTyryPheThrGlyPheGlyLysHisGlyAspThrAlaValArgGlyLysHisAsp 349
Db 7409 TTCTACTTACCTGCGCTTGTGTGTGATGAAGATGCGGACTTCTGGCAAGATTCGAT 7468
QY 350 ProAlaTyryMetValHisAspPheGlnLeuMetLysTrpIleGlyAlaAsnSerPheArg 369
Db 7469 AACGTGCTGATGTGTGACACACACGATTAATGATGTGATGGGCCAATCTCTACCT 7528
QY 370 ThrSerHisTyryProTyryAlaGluGluValMetAspPheAlaAspArgAsnGlyIleVal 389
Db 7529 ACTCGCATTAACCTTACCTGACAGATGATCGTGCAGATGATGATGATGATGATGATG 7588
QY 390 ValIleAspGlnThrProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGlu 408
Db 7589 GTGATGATGAAGACTGCTGCTGCTGCTTAACTCTCTTAAAGCATGTGTTCCAGAGC 7648
QY 409 SerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAla 427
Db 7649 GGCACAAAGCCGAAAGAAAGTGAACGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7708
QY 428 HisLysGlnAlaIleArgGluLeuIleAlaGlyAspLysAsnHisAlaSerValValMet 447
Db 7709 CACTTACAGGCGATTAAGAGCTGATAGCGCGTGAACAAACCAACCAACGCGTGTGATG 7768
QY 448 TrpSerIleAlaAsnGluProAlaSerHisGlyAspGlyValArgGlyLysTyryPheGluPro 467
Db 7769 TGGAGTATGTCACCAACACCGGATACCCGTCGCAAGTGTGCAAGGAAATTTCCGCCCA 7828
QY 468 LeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGly 487
Db 7829 CTGGCGGAAAGCAACGGGTAACTCGACCGGACCGGATCACTGCGCTCAATGTAAATG 7888
QY 488 ThrAlaThrTyryGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArg 507
Db 7889 TTCTGACAGCTTACACCATATACCATACAGCATCTTGTGATGTGTGTGCTGCAACCGT 7948
QY 508 TyryPheGlyTrpTyrySerGlnThrGlyAspLeuGluGluAlaGluAlaIleGluGlyLys 527
Db 7949 TATTTACGATGGATGTCCAAAGCGGATTTGAAACCGGACAGAAAGTACTCGGAAAAA 8008
QY 528 GlnLeuHisGlyTyryGlnGluLysPheHisArgProIleValMetThrGluTyryAla 547
```

```

|||||
Db      GAACCTTGCGCTGCGAGGAGAACTGCATCAGCCGATTATCATCACCAGATACGGCGTG 8068
Qy      AspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyrSerGluGluPheGlnVal 567
Db      548 AspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyrSerGluGluPheGlnVal 567
Qy      8069 GATACGTTAGCCGGCTGCATCATGATACACCCAGCATGTGTGAAGATATCATCAGTGT 8128
Qy      568 GlnMetLeuAspMetTyrHisArgValPheAspArgIleGluSerMetAlaGlyGlnHis 587
Db      8129 GCATGGCTGATATGATACACCGCTCTTTCATGCGCTCAGCCGCGTGTGGAACAG 8188
Qy      588 ValTyrAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnLys 607
Db      8189 GATGGAATTTCCGCGATTGTCAGCTCCGACGAGCATATTGCGCTTGGCGGTAAACAG 8248
Qy      608 LysGlyValPheThrArgAspArgLysProLysAlaAlaHisSerLeuArgAlaArg 627
Db      8249 AAGGAGATCTTCACTCCGACCGCAACCGAAGTGGCGGCTTTCTGTCGCAAAAAGC 8308
Qy      628 TyrThrSerIleAsp 632
Db      8309 TGGACTGCGATGAAC 8323

```

```

RESULT 11
US-09-845-064-50
; Sequence 50, Application US/09845064
; Publication No. US20030175976A1
; GENERAL INFORMATION:
; APPLICANT: MERISTEM THERAPEUTICS
; TITLE OF INVENTION: CLEAN SYNTHETIC VECTORS, PLASMIDS, TRANSGENIC PLANTS
; TITLE OF INVENTION: AND PLANT PARTS CONTAINING SAID VECTORS, AND THEIR
; FILE REFERENCE: SYNVECI
; CURRENT APPLICATION NUMBER: US/09/845,064
; CURRENT FILING DATE: 2001-04-27
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 50
; LENGTH: 9688
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: pMRT1334
; NAME/KEY: misc_feature
; LOCATION: (1)
; OTHER INFORMATION: pMRT1334 was obtained by replacing the mpIII
; OTHER INFORMATION: expression cassette of pMRT1206 by the mpIII
; OTHER INFORMATION: expression cassette of pBIN19
US-09-845-064-50

```

```

Alignment Scores:
Pred. No.: 1.04e-168 Length: 9688
Score: 1714.50 Matches: 337
Percent Similarity: 69.12% Conservative: 95
Best Local Similarity: 53.92% Mismatches: 177
Query Match: 51.12% Indels: 17
DB: 10 Gaps: 9

```

```

US-10-757-093-4 (1-634) x US-09-845-064-50 (1-9688)
Qy      15 ProSerIleuThrProAlaAlaArgHisPheProArgAsnGluMetThrGlnHisGlu 34
Db      6776 CCCAAGCTTGGCCGCGCGCTTAACACGCGTGATCTT--ATTTAAGTCGACTAGGGGT 6833
Qy      35 GlnProLeuIleuValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAsp 54
Db      6834 GGTACAGCCCTTAAGTACGCTGTAGAAACCCCAACCCGTGAATCAAAAAAAGTCGAC 6893
Qy      55 GlyLeuTyrLysPheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnPro 71
Db      6894 GGCCCTGTGGCATATTCAGTCTGGATCGGAAAAAAGTGTGAATGTATCAG-----CGTTGG 6947

```

```

Qy      72 TyrThrAlaProLeuProLysGlyLeuGluCysProValProAlaSerTyrAsnAspIle 91
Db      6948 TGGGAAAGCCCGTTACAGAAAGCCGGCAATTGCTGCGACGAGCTTTAAACGATCAG 7007
Qy      92 PheIleSerArgGluIleHisAspHisValGlyTyrValTyrTyrGlnArgGluValIle 111
Db      7008 TTCGCCGATGCACATATTCCTAATTTATGCGCGCAACCTCTGATATCATCAGCCGAAAGTCTTT 7067
Qy      112 ValProLysGlyTyrPheSerGlnArgTyrLeuValArgAlaGluSerAlaThrHisHis 131
Db      7068 ATACCGAAAGCTTGGCGAGCCGATGCTGCTGCTGCTTCATTCAGTGCATCATTTAC 7127
Qy      132 GlyArgIleTyrValAsnAsnArgLeuValAlaGlyIleValGlyTyrThrProPhe 151
Db      7128 GCGAAAGTGGGATCAATATACAGAAAGTATGAGAGATAGAGGCGGCTATACCCCATTT 7187
Qy      152 GluAlaAspValThrGluLeuValAlaProGlyGluLysPheAspArgLeuThrIleGlyVal 171
Db      7188 GAAGCCGATGTCAAGCGCGTATTTATTTGCGGGAAGATGTACGTTACACCGTTTGTGTG 7247
Qy      172 AsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThr 191
Db      7248 AACACGACACTGAACCTGGCAGACTATCCCGCGGAATGGTGTATACC--GACGAAAC 7304
Qy      192 GlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSer 211
Db      7305 GCGAAGAAAGAGCTTATCTTCATGATTTCTTATACATACCGGAATCCATCCGACG 7364
Qy      212 IleTyrPheLysSerValProGlnGlnHisIleGlnAspIleThrValValThrAspVal 231
Db      7365 GATAGTCTTACACACACGCGAAGACACCTGGTGGACATATCACCGGTGAGCAGCATGTC 7424
Qy      232 -----AspGlyAspAsnGlyLeuIleAsnTyrGluValGluValAlaAsnGlnThr 249
Db      7425 GCGCAAGACTGTATCACACCGCTGTGACTGCGACGCTG--GTGGCAAT----- 7472
Qy      250 GlyGlnIleGlnIleSerValIleAspGluAspGlyValAlaIleValAlaLysSerGly 269
Db      7473 GTGATGTCAGCGTTGATGACGCTGATGCGGATCAACAGGTGTTGCACTGCAACAAAGC 7532
Qy      270 AlaglnGlyThrValThrIleProSerValLysLeuTyrGlnProGlyAlaAlaTyrLeu 289
Db      7533 ACTAGCGGACCTTGGCAAGGTGAATCCGCACTCGCAACCGGAGTGAAGGTATCTC 7592
Qy      290 TyrGlnLeuGlnValAsnIleValIleGlySerSerGlyAspValValAspThrTyrAsnLeu 309
Db      7593 TATGAACCTGTGCGTC-----ACAGCCMAAAGCCAGACAGATGTATCTACCCGCTT 7646
Qy      310 AlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysPro 329
Db      7647 CGCGTCGACATCCGATGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7706
Qy      330 PheTyrPheThrGlyPheGlyLysHisGluAspThrAlaValArgGlyLysGlnHisAsp 349
Db      7707 TTCTACTTATCTGCGCTTGTGCTGATGAAGATCGGAGATCGGTGCGCAAAAGATTCGAT 7766
Qy      350 ProAlaTyrMetValHisAspPheGlnLeuMetLysTyrIleGlyValAsnSerPheArg 369
Db      7767 AACGTGTATGTGTGACGACGACGACGATTAATGAGCTGATGAGGAGCAACTCTACCGT 7826
Qy      370 ThrSerHisTyrProTyrArgIleGluGluValMetAspPheAlaAspArgAsnGlyIleVal 389
Db      7827 ACCCTGCAATTAACCTTAACGCTGAAGAGATGCTGATCGGCAATGAACATGAGCATCGTG 7886
Qy      390 ValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGlu 408
Db      7887 GTATGATGAATACTGCTGCTGCGCTTAACTCTCTTAAAGCAATGGATTGCAAGCG 7946
Qy      409 SerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAla 427
Db      7947 GGCACAAAGCCGAAAGAACTGTACAGGAAAGCATCAACCGGAAACTCAGCAAGCG 8006
Qy      428 HisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMet 447

```



```

Db      8549 TTCTACTTACTGCTTGGTCGTCATGAAAGATCGGACCTGGTCGCAAGATTCGAT 8490
Qy      350 ProAlaTyrMetValHisAspPheGlnLeuMetLysTrpIleGlyAlaAsnSerPheArg 369
Db      8489 AACGTCTGATGGTGCACGACCGCATTAATGACATGAGTGGGGCAACTCTCTACCGT 8430
Qy      370 ThrSerIleTyrProTyrAlaGluGluValMetAspPheAlaAspArgAsnGlyIleVal 389
Db      8429 ACCCTCCGATTACCTTACGCTGAGAGATGCTCGACATGGGCAATGCAATGGCATGCTG 8370
Qy      390 ValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu--MetGlyValSerGlu 408
Db      8369 GTGATTCATGAACCTGCTGCTGCTTAACTCTTAACTCTTAACTCTTAACTCTTAACTCT 8310
Qy      409 SerGlyAlaPro--GlnThrPheThrProAspAlaIleAsnAspLysTrpGlnGluAla 427
Db      8309 GGCACACAGCCGAGAAAGCTGACAGCGAAGAGCACTCAACGGGGAAGCTCAGCAAGCG 8250
Qy      428 HisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMet 447
Db      8249 CACTTACAGCGCATTTAAGAGCTGATAGCGGTGACAAAACCAACCAAGCGTGGTATG 8190
Qy      448 TrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGluTyrPheGluPro 467
Db      8189 TGGAGTATTTGCCAAGCAAGCGGATACCCGTCGCGAAGTGCACGGAAATTTTCGCCCA 8130
Qy      468 LeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGly 487
Db      8129 CTGCGCGAGCAACCGCTAACTGACCGACCGCTCGATTCACCTCGCATATGTATG 8070
Qy      488 ThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArg 507
Db      8069 TTCTGCCACGCTACACCGAATACCATGAGCATCTTTAATGTGTCTGTCTGTGAACCGT 8010
Qy      508 TyrPheGlyTyrPyrSerGlnThrGlyAspLeuGluGluAlaGluAlaIleGluLys 527
Db      8009 TATTTACGATGATATGTCGAAGCGGATTTGGAAACGGCAGAGAACTGGAATAA 7950
Qy      528 GluLeuHisGlyTyrGlnGluLysPheHisArgProIleValMetThrGluTyrGlyAla 547
Db      7949 GAATCTTCTGCGCTGCGAGGAAACTGCATCAGCCGATTCATCACCAGATTCGCGCTG 7890
Qy      548 AspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProThrSerGluGluPheGlnVal 567
Db      7889 GATACGTACCGCGGCTGCACTCATATTCACCGACATGTGAGTGAAGATATCACTGT 7830
Qy      568 GluMetLeuAspMetTyrHisArgValPheAspArgIleGluSerMetAlaGlyLysHis 587
Db      7829 GCATGCTGATATGATTCACCGGCTTTGATGCTGCTCAGCCGCTCGGTGAACAG 7770
Qy      588 ValTrpAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnLys 607
Db      7769 GATTCGAAATTCGCCGATTTTGGCACTTCGCAAGGCAATATTTGCGCGTTCGCGTAACAG 7710
Qy      608 LysGlyValPheThrArgAspArgLysProLysAlaIleAlaHisSerLeuAlaArg 627
Db      7709 AAAGGAGTCTTCACTCCGACCGCAACCGAAGTGGCGGCTTTTCTGTCGCAAAAACGC 7650
Qy      628 TrpThrSerIleAsp 632
Db      7649 TGGACTGCGCATGAAC 7635

```

```

; FILE REFERENCE: 70149USNP
; CURRENT APPLICATION NUMBER: US/10/680,824A
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/418596
; PRIOR FILING DATE: 2002-07-10
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 10011
; TYPE: DNA
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: plasmid pEB10
; US-10-680-824A-19

Alignment Scores:
Pred. No.: 1,23e-168 Length: 10011
Score: 1714.00 Matches: 331
Percent Similarity: 70.50% Conservative: 92
Best Local Similarity: 55.17% Mismatches: 161
Query Match: 51.10% Indels: 16
DB: 18 Gaps: 9

US-10-757-093-4 (1-634) x US-10-680-824A-19 (1-10011)
Qy      40 ValArgProGlnArgHisSerArgGluLeuValAsnLeuAspGlyLeuTrpLysPhe 59
Db      2073 GTCCGCTCTGTAGAACCCCAACCGGAAATGAAATCAAAACCGCCCTGTGGCATTC 2014
Qy      60 AlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProThrAlaProLeu 76
Db      2013 AGCTGTGATGCGAANAACGTGGAATTCATCAG-----CGTTGGTGGGAACCGCTTA 1960
Qy      77 ProLysGlyLeuGluCysProValProAlaSerTyrAsnAspIlePheIleSerArgGlu 96
Db      1959 CAAGAAAGCCGGCGCAATTCGCTGCCAGCGCATTTTAACGATCATTCGCGATGCAAGAT 1900
Qy      97 IleHisAspHisValGlyTyrValTyrTyrGlnArgGluValIleValProLysGlyTrp 116
Db      1899 ATTGTAATTAATCGGAGCAACGCTGTGTATCAGCGGAAAGTCTTTATACCGAAAGTTGG 1840
Qy      117 SerGlnGluArgTyrLeuValArgAlaGluSerAlaThrHisIleGlyArgIleTyrVal 136
Db      1839 GCAGGCGAGGTAATCGTGTGCTGCTTCGATCGCGTCACTCATTAACGCAAAAGTGGTTC 1780
Qy      137 AsnAsnArgLeuValAlaGluHisValGlyGlyTyrThrProPheGluAlaAspValThr 156
Db      1779 AATTAATCAGAAAGTATGAGCATTCAGGCGCGCTATTAACGCTTGAAGCCGATGCACG 1720
Qy      157 GluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyValIleAsnGluLeuThr 176
Db      1719 CCGTATGTTATTTCCGAGAAAGTGAATGATACCGTTTGTGTGAACAAGACTGAAC 1660
Qy      177 HisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysArgIleGln 196
Db      1659 TGGCAACATATCCCGCGGAAATGTGATTAAC---GACGAAACCGCAAGAAAGACAG 1603
Qy      197 ThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIleTrpLeuTyrSer 216
Db      1602 TCTTACTTCATGATTTCTTAACTATGCGGGATCATTCGAGCGTAATGCTCTACAC 1543
Qy      217 ValProGlnGlnHisIleGlnAspIleThrValValThrAspVal-----AspGlyAsp 234
Db      1542 ACCGCCAACAACCTGGGTGAGCATATCACCGGTGTACCGATGTCGCGCAAGCTGTAAAC 1483
Qy      235 AsnGlyLeuIleAsnTyrGluValAlaAsnGlnThrThrGlyGlnIleGlnIle 254
Db      1482 CAGCGCTGTGTACTGCGCAGGTG---GTGGCCAAI-----GCTATGTCAACGTT 1435
Qy      255 SerValIleAspGluAspGlyAlaIleValAlaLysAlaSerGlyAlaGlnGlyThrVal 274
Db      1434 GAATCGCGATGCGGATCAACAGTGTGTCAACTGCAACAGGCACTAGCGGACCTTTG 1375

```

```

; RESULT 13
; US-10-680-824A-19/c
; Sequence 19, Application US/10680824A
; Publication No. US20040133937A1
; GENERAL INFORMATION:
; APPLICANT: Boudreau, Eric
; APPLICANT: Gu, Weidong
; APPLICANT: De Fremond, Anic
; APPLICANT: Heifetz, Peter
; TITLE OF INVENTION: Plasmid Transformation

```

Qy	275	ThrIleProSerValLyLeuTrpGlnProGlyAlaIaIaTyrlenTyGlnLeuLlnVal	294
Db	1374	CAAGTGGTAATCCGCACCTCTGGACACCGGGTGAAGGTATCTTAAGAACTGGCGTC	1311
Qy	295	AsnIleValGlySerSerGlyAspValValAspThrTyAsnLeuAlaThrGlyValAlaG	314
Db	1314	-----ACAGCGAAAGCCAGACAGAGTGTGATATCTAACCGCGCTTCGGGTGGCATCCGG	126
Qy	315	ThrValLySValAlaGlySerGlnPheLeuIleasnGlyLySProPheTyPheThrGly	334
Db	1260	TCAGTGGCAGTGAAGGGGGCAACAGTCTCTGATTAACCAACAAACCTTCTACTTAACTGGC	1203
Qy	335	PheGlyLySHisGlyAspThrAlaValArgGlyLySGLYHisAspProAlaTyMetVal	354
Db	1200	TTTGGTCTCATGAAGATGCGGACCTGGTGGTGGCAAAAGATTCGATTAACGTCTGATGGTG	114
Qy	355	HisAspPheGlnLeuMetLysTrpIleGlyAlaAsnSerPheArgThrSerHisTyPro	374
Db	1140	CACAGCACCGCATTAATGACTGCGATTGGGGCAACTCCTACCGTAACCTCGCATTAACCT	108
Qy	375	TyrAlaGlnGlyValMetAspPheAlaAspArgasnGlyIleValValIleAspGluThr	394
Db	1080	TACGCTGAAGAGATGCTGCACCTGGCGCAATGAACATGGCATGTGTGATGATGAACCT	1022
Qy	395	ProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGluSerGlyAlaPro---	412
Db	1020	GCTGTGTGCGCTTTAACTCTCTTTAGCATGTGTTCGAACGGCGCAACAGCGAA	961
Qy	413	GlnTrpPheTrpProAspAlaIleAsnAspLySThrGlnGluAlaHisGlySlnAlaIle	432
Db	960	GAACTGTACAGCAAGAGGACAGTCAACGGGAAACTCAGCAAGCCCATTAACGGCGATT	901
Qy	433	ArgGluLeuIleAlaArgAspLySAsnHisAlaSerValValMetTrpSerIleAlaAsn	452
Db	900	AAAGAGCTGATGACCGCTGCACAAAACCAACCCACGCTGATGTGAGTTCCTCCAAC	841
Qy	453	GluProAlaSerHisGlyAspGlyAlaArgGlyTyrlenGlyProLeuThrAsnLeuThr	472
Db	840	GAACCGGATACCCCGTCCGCAAGSTGCACGGGAATATTCGCGCACTGGCGGAACGACAG	781
Qy	473	ArgGlnLeuAspProThrArgProIleTrpPheAlaAsnValGlyThrAlaThrTyGln	492
Db	780	CGTAACTCGACCCCAACGCGTCCGATCACTGCGTCAAGTAAATGTTCTGGCAAGCTCAC	721
Qy	493	LeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrlenPheGlyTyrlen	512
Db	720	ACCGATACCATCAGCGATCTTTGATGTGCTGCGCTGAACCGTTATTAACGATGGTAT	661
Qy	513	SerGlnThrGlyAspLeuGlnGluAlaGlyAlaAlaLeuGlnLySLeuHisGlyTrp	532
Db	660	GTCCAACACGGGATTTGGAAACGCGACAGAGAGTACTGGAAAAAGAACTTCTGGCGTGG	601
Qy	533	GlnGlyLySPhenHisArgProIleValMetThrGlnTyrlenGlyAlaAspThrLeuAlaGly	552
Db	600	CAGGAGAAACTCACTACCGCATTTATCATACCGAATACGGCGTGATACGTTAGCCGGG	541
Qy	553	LeuHisSerIleLeuGlyLeuProTrpSerGlnGluPheGlnValGlnMetLeuAspMet	572
Db	540	CTGCACCTCAATGTACACCGACATGTGGAAGTAAAGTATCAAGTGTACAGCGCTGATATG	481
Qy	573	TyrlenHisArgValPheAspArgIleGlnSerMetAlaGlyGlnHisValIleTrpAsnPheAla	592
Db	480	TATCACCGGCTTTGATCGCGTCCAGCGCGCTGTCGGTGAACAAGGTATGAATTTCCGC	421
Qy	593	AspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnLySLeuGlyValPheThr	612
Db	420	GATTTTGCACCTCGACAGCAATATGGCCGTTGGCGGTAACAAGAAAGGATCTTCACT	361
Qy	613	ArgAspArgLySProLySAlaAlaAlaHisSerLeuArgAlaArgTrpThrSerIleAsp	632
Db	360	CGCGAACCGCAACCGAAGTCGGCGGCTTTCTGCTGTCAAAAACGCTGACCTGCATGAAC	301

```

US-10-336-566-83
; Sequence 83. Application US/10336566
; Publication No. US20030175292A1
; GENERAL INFORMATION:
; APPLICANT: Robinson, Harriet L.
; APPLICANT: Smith, James M.
; APPLICANT: Hua, Jian
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING
; FILE REFERENCE: 12804-006001
; CURRENT APPLICATION NUMBER: US/10/336,566
; PCT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: US 10/093,953
; PRIOR FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 09/799,675
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: PCT/US01/06795
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 60/251,083
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 60/186,364
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/324,845
; PRIOR FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 83
; LENGTH: 12224
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetically generated plasmid sequence-pLW-48
US-10-336-566-83

Alignment Scores:
Pred. No.:          1.63e-168      Length:      12224
Score:              1714.00        Matches:     331
Percent Similarity: 70.50%         Conservative: 92
Best Local Similarity: 55.17%       Mismatches:  161
Query Match:         51.10%         Indels:      16
DB:                  16             Gaps:        9

US-10-757-093-4 (1-634) x US-10-336-566-83 (1-12224)

QY      40 ValArgProGlnArgThrSerSerArgGluValAsnLeuArgGlyLeuTrpLyPhe    59
      ||| ||||| ||||| ||||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      979 GTAGCGTCCTGTAGAAACCCCAACCCTGGAAATCAAAAAACCTGCAGCGCCTGTGGCATTC   1038
QY      60 AlAleu-----AlaserGlyLeuAsnAspThrIagInProTrrPTThAlProleu    76
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1039 AGTCGTGATGCCGAAMAACCTGTGGATGATGATAG-----CGTTGGTGGAAGCGCGTTA   10922
QY      77 ProLysgGlyLeuGlucyProValProAlasertyrAsnAspIlePhelIeserArgGlu    96
      :::: |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db      1093 CAAGAAGCCGGCGCAATTGCTGTCGACGGCGATTTAACGATCAGTTCGCCGATGCAGAT   11522
QY      97 IlehIsaphtIsValGYtTPvalTYtytYSlnArgGluValIleValProLygLYtTP    116
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1153 AYTGGTAATTATGCGGGCAACGTCGTGGATCAGCGCAAGCTTATAACCAAGATTGG    12122
QY      117 serGlnGuArGyrlyLeuValArgAlaGluSerAlaThrshIShISglYArgIletyrVal    136
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1213 GCAGGCGCGCGTATCGTCGCTGCGTTTCGATCCGCTACTATVACGCCMAAGTGTGGTTC   12727
QY      137 AsnAsnArgyleValAlaGluHISValIGlyTYtThrProPheGluValaASPValThr    156
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1273 AAATAATCGAAGATGATGAGACATCAGGGCGGCTATACGCCATTGGAAGCCGATGTCAG   13333
QY      157 GlueValAlaAProGlyGlyLysPheArgLeuThrIleGlyValaAsnAsnGlyLeuThr    176
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      1333 CCGATATGTTATTCGCCGGAAAAGTACGATCATCCGTTGTGTGAACAAGAACCTGTGAC    13992
QY      177 HisGluThrIlePProProGlyLysIleThrThrGlyAsnAlaThrnglyLyArgIleIn    196

```



```

Db      1393 TGCAGACTATCCGCGCGGAATGTGATTAC---ACAGAAACGGCAAGAAACAG 1449
Qy      197  ThTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIleTyrPheTyrSer 216
Db      1450 TCTTACTTCATGATTTCTTAACTATGCGGAATCCATGCGACCGTAATGCTCTTACACC 1509
Qy      217  ValProGlnHisIleGlnAspIleThrValValThrAspVal-----AspGlyAsp 234
Db      1510 ACCCGCAACACCTGGGCGAGATATCACCGTGTGACGATGCGCGCAAGACTGTAAAC 1569
Qy      235  AsnGlyLeuIleAsnTyrGluValAlaGlnGlnThrGlyGlnIleGlnIle 254
Db      1570 CACGCGCTGTGACTGCGCAGGTG---GTGGCCAAAT-----GGTGAATGCGCGGT 1617
Qy      255  SerValIleAspGluAspGlyAlaIleValAlaValAspGlyAlaGlnGlyThrVal 274
Db      1618 GAACCTGGTATGGGATTCACACAGTGTGCACTGAGACAAAGCACTAGCGGACTTTG 1677
Qy      275  ThrIleProSerValIleAsnTyrGlnProGlyValAlaIleTyrLeuTyrGlnLeuGlnVal 294
Db      1678 CAAGTGTGATTCGCGACCTCTGTGCAACCGGTGACAGTATCTCTATGAACTGTGCTC 1737
Qy      295  AsnIleValGlySerSerGlyAspValValAspThrTyrAsnLeuAlaThrGlyValArg 314
Db      1738 -----ACAGCCAAAGCCAGACAGAGTGTGATCTACCGCGCTGCGTCCGCGCATCCG 1791
Qy      315  ThrValIleValAlaGlySerGlnPheLeuIleAsnGlyIleProPheTyrPheThrGly 334
Db      1792 TCAGTGTGACATGAAAGGCGGACAGTCTCTATTAACCAACAAACCGTTCTCTTACGCGC 1851
Qy      335  PheGlyIleHisGlnAspThrAlaValArgGlyIleAspGlyHisAspProAlaTyrMetVal 354
Db      1852 TTTGGTGTATGAAAGTGGCGACTTCGTGGCAAGATTCGATTAACGTGCTGATGATG 1911
Qy      355  HisAspPheGlnLeuMetIleTyrIleGlyAlaAsnSerPheArgThrSerHisTyrPro 374
Db      1912 CACGACCAACCATTAATGATGATGATGGGCGCAATCTCAACGCTACCTCGCATTAACCT 1971
Qy      375  TyrAlaGlnGluValMetLeuPheAlaAspArgAsnGlyIleValIleAspGlnThr 394
Db      1972 TAGCGTGAAGATGCTCGACTGGCAGATGAATGCAATCGGTGATGATGAAACT 2031
Qy      395  ProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGluSerGlyAlaPro--- 412
Db      2032 GCGCTGTGCGCTTTAACTCTCTTGAAGCTGTTCCGAAGCGGCAACAGCCGAAA 2091
Qy      413  GlnThrPheThrProAspAlaIleAsnAspIleThrGlnGluAlaHisIleValIle 432
Db      2092 GAACTGTACAGCGAAGGCGACGTCAACGCGGAAACTCAGCAAGCGCACTTACAGCGCAT 2151
Qy      433  ArgGlyLeuIleAlaIleAspIleValAsnIleAspValIleMetPheSerIleAlaAsn 452
Db      2152 AAAGAGCTGTAGGCGCTGACAAACACCAAGCCTGTGATGTGAGATATGGCCAAC 2211
Qy      453  GlnProAlaSerHisGlnAspGlyAlaArgGlnTyrPheGluProLeuThrAsnLeuThr 472
Db      2212 GAACCGGATACCGCTCGCAAGGCGAGGAAATTTCCGCGCACTGCGGGAAGCAAG 2271
Qy      473  ArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyrGln 492
Db      2272 CGTAACTGACCGACCGCGCTCGCATCAGCTGCAATGTAATGTTCTCGACGCTCAC 2331
Qy      493  LeuAspArgIleSerAspLeuPheAspValSerCysAlleAsnArgTyrPheGlyTyr 512
Db      2332 ACCGATTAACCTACGCGATCTCTTGAATGCTGTGCTGTAACCGTTTACCGATGTAT 2391
Qy      513  SerGlnThrGlyAspLeuGlnGluAlaGluAlaIleAsnGlyIleValIleValIle 532
Db      2392 GTCCAAAGCGCGATTTGAAACGCGAGAGATCTGAAAGAAAGAACTTTGCGCTGG 2451
Qy      533  GlnGluIlePheHisArgProIleValMetThrGluTyrGlyAlaAspThrLeuAlaGly 552

```

```

Db      2452 CAGAGAACTGATCAGCCGATTAATCATCCGAATACCGCGCTGTGATACCGCGG 2511
Qy      553  LeuHisSerIleLeuGlyLeuProTyrSerGluGluPheGlnValGlnMetLeuAspMet 572
Db      2512 CTGACTCAATATGATACCCAGCAGATGTGAGTGAAGATGATAGTGTGATGCGATGATG 2571
Qy      573  TyrHisArgValPheAspArgIleGluSerMetAlaGlyGluHisValTyrAsnPheAla 592
Db      2572 TATCACCGCGCTTTGATGCGCGTCAAGCGCGTGTGCGTGAACAGTATGGAATTTGCC 2631
Qy      593  AspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnIleValIlePheThr 612
Db      2632 GATTTTGCAGCTTCGCAAGGCAATATGCGCGTGGGTAAACAAAGAAAGGATCTTCACT 2691
Qy      613  ArgAspArgIleProIleAlaIleAlaHisSerLeuArgAlaArgTyrThrSerIleAsp 632
Db      2692 CCGGACCGCAACCGAAGTGTGCGCGCTTTCTGTCTGCAAAACGCTGAGCATGCAAC 2751

RESULT 15
US-10-424-638-1/c
; Sequence 1, Application US/10424638
; Publication No. US20030175245A1
; GENERAL INFORMATION:
; APPLICANT: BROUGH, DOUGLAS E
; APPLICANT: KING, C R
; APPLICANT: KOVESDI, IMRE
; APPLICANT: SCHAIKLE, JASPER J
; TITLE OR INVENTION: REPLICATION DEFICIENT ADENOVIRAL TNF VECTOR
; FILE REFERENCE: 222226
; CURRENT APPLICATION NUMBER: US/10/424, 638
; PRIORITY FILING DATE: 2003-04-25
; PRIOR APPLICATION NUMBER: US 09/604,694
; PRIORITY FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 32798
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
; US-10-424-638-1

Alignment Scores:
Pred. No.: 6,64e-168 Length: 32798
Score: 1714.00 Matches: 331
Percent Similarity: 70.50% Conservative: 92
Best Local Similarity: 55.17% Mismatches: 161
Query Match: 51.10% Indels: 16
Gaps: 9

US-10-757-093-4 (1-634) x US-10-424-638-1 (1-32798)
Qy      40  ValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTyrPhe 59
Db      31982 GTCCGCTCTGTAAACCCAGCCGTAATCAAAAACTGACGCGCTGTGGCATTC 31923
Qy      60  AlLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTyrThrAlaProLeu 76
Db      31922 AGCTGTGATCGCAAAACTGTGGAATGTATCAG-----CGTTGCTGCGAAGACGCGCTTA 31869
Qy      77  ProIleGlyLeuIleCysProValProAlaSerTyrAsnAspIlePheIleSerArgGlu 96
Db      31868 CAAAGAAAGCCGGCAATGCTGTGCGAGCGAGTATTAAACGATCAGTTCGCCGATGCAGAT 31809
Qy      97  IleHisAspHisValGlyTyrValTyrTyrGlnArgGluValIleValProIleArgIleTyr 116
Db      31808 ATTGTAATTAATGCGGCAACGCTGTGATCAACGCGGAAGTCTTTATCCGAAAGGTGG 31749
Qy      117  SerGlnGluArgTyrIleValAlaGluSerAlaThrHisHisGlyArgIleTyrVal 136
Db      31748 GCAGGCGAGGATGCTGTGCGCTTTCGATGCGGTGATCATCATTAACGCAAAAGTGTGGTCC 31689

```

QY 137 AsnAsnArgLeuValAlaGluHisValGlyGlyTyrThr-ProPheGluAlaAspValThr 156  
 Db 31688 AATAATCAGGAAGTGTATGAGCATCAGGGCGGCTATACCCATTGTAACCCAGTGCACG 31629  
 QY 157 GluLeuValAlaAspGlyGlyLysPheArgLeuThrIleGlyValAlaAsnGluLeuThr 176  
 Db 31628 CCGTATGTTATTCGCCGGAAGTACGTATACCCGTTGTGTGAACAACAACTGAAC 31569  
 QY 177 HisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysArgIleGln 196  
 Db 31568 TGGCAGCATATCCCGCGGGAATGTGATTACC--GACGAAAAACGCAAGAAAAAGCAG 31512  
 QY 197 ThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIleTyrLeuTyrSer 216  
 Db 31511 TCTTACTTCATATTTCTTAACTTACCGGAATCCATCGCAGCGTAATGCTTACACC 31452  
 QY 217 ValProGlnGlnHisIleGlnAspIleThrValIleThrAspVal-----AspGlyAsp 234  
 Db 31451 ACCGCCAACCCTCGGTGAGCATATCACCGTGTACCGCATGTCCGCAAGACTGTAAAC 31392  
 QY 235 AsnGlyLeuIleAsnTyrGluValGluValAlaAsnGlnThrThrGlyGlnIleGlnIle 254  
 Db 31391 CACGCGCTGTGTACTGTCAGGTG---GTGGCCAAAT-----GGTGATGTCACCGGT 31344  
 QY 255 SerValIleAspGluAspGlyAlaIleValAlaLysAlaSerGlyAlaGlnIleThrVal 274  
 Db 31343 GAACCTGGTATGGATGATCAACAGGTGTGCACTGGACAAAGCACTAGCGGGAATTTTG 31284  
 QY 275 ThrIleProSerValLysLeuTyrGlnProGlyValAlaIleTyrLeuTyrGlnLeuGlnVal 294  
 Db 31283 CAGTGTGTGAATCCGCACTCTGCGCAACCGGTGAAGATTATCTATGAACTGTGCGTC 31224  
 QY 295 AsnIleValGlySerSerGlyAspValValAspThrTyrAsnLeuAlaThrGlyValArg 314  
 Db 31223 -----ACAGCCAAAGCCAGACAGAGTGTATATCAACCGCTTCGCGTGGCATCCGG 31170  
 QY 315 ThrValIleValAlaGlySerGlnPheLeuIleAsnGlyLysProPheTyrPheThrGly 334  
 Db 31169 TCGATGGCAGTGAAGGCGCAAGCTTCTGATTACCAACAACCGTTCTTACTGCGC 31110  
 QY 335 PheGlyLysHisGlyAspThrAlaValArgGlyLysGlyHisAspProAlaTyrMetVal 354  
 Db 31109 TTTGCTGCTATGAAGATGCGGACTGCTGCGCAAGATTGATACGTGCTGATGTG 31050  
 QY 355 HisAspPheGlnLeuMetLysTyrIleGlyAlaAsnSerPheArgThrSerHisTyrPro 374  
 Db 31049 CACGACCAACGATTAATGACTGTGATTGGCGCAACTCTACCGTACTCGCATTTACCT 30990  
 QY 375 TyrAlaGluGluValMetAspPheAlaAspArgAsnGlyIleValAlaIleAspGluThr 394  
 Db 30989 TACGCTGAAGATGCTCGACTGGGCAAGTGAACATGCGATCGTGTGATTGAAACT 30930  
 QY 395 ProAlaValGlyLeuAsnIleAlaLeu--MetGlyValSerGluSerGlyAlaPro--- 412  
 Db 30929 GCTGCTGCGGCTTTAACTCTCTTTAGGCATGTGTTTCGAAGCGGCAACAAAGCCGAAA 30870  
 QY 413 GlnThrThrThrProAlaAlaIleAsnAspLysThrGlnGluAlaHisLysGlnAlaIle 432  
 Db 30869 GAACGTGACAGCGAAGCGTCAACGGGAAACTCAAGCAAGCGCACTTACAGCGCAATT 30810  
 QY 433 ArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMetTyrSerIleAlaAsn 452  
 Db 30809 AAAAGCTGATAGCGCGTGAACAAAACCAACCAAGCGGTGATGTGAAGTATGGCAAC 30750  
 QY 453 GluProAlaSerHisGluAspGlyAlaArgGlyLysTyrPheGluProLeuThrAsnLeuThr 472  
 Db 30749 GAACCGGATACCCGTCGCAAGTGCACGGGAATATTTCCGCCACTGGCGGAAGCAACG 30690  
 QY 473 ArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyrGln 492  
 Db 30689 CGTAAACTCGAACCAGCGCTCGATCACCTGCGCTCAATGTAAATGTTCTGCGACGCTCAC 30630  
 QY 493 LeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTyr 512

Db 30629 ACCGATACCATCAGCGATCTCTTATGTCGTGCTGAAACCGTTATTACGATGTAT 30570  
 QY 513 SerGlnThrGlyAspLeuGluGluAlaGluAlaLeuGluLysGluLeuHisGlyTyr 532  
 Db 30569 GTCCAAAGCGGCGATTGGAAAACGGCAGAGAAGTACTGGAAAAAGAACTTGGCCCTGG 30510  
 QY 533 GlnGluLysPheHisArgProIleValMetThrGluTyrGlyAlaAspThrLeuAlaGly 552  
 Db 30509 CAGGAAACCTGCATCAGCCGATTATCATCACCGAAATACGGCGTGAATACGTTAGCCGG 30450  
 QY 553 LeuHisSerIleLeuGlyLeuProTyrPheSerGluGluPheGlnValGlnMetLeuAspMet 572  
 Db 30449 CTGCACTGAATGTACACCGACATGTGAGTGAAGATACAGTGTGATGCTGATATG 30390  
 QY 573 TyrHisArgValPheAspArgIleGluSerMetAlaGlyGlnHisValThrAsnPheAla 592  
 Db 30389 TATCACCGCGCTTGTGATCGCGTCAAGCCGCTGCGGTGAACAGGTATGGAAATTCGCC 30330  
 QY 593 AspPheGlnThrAsnLeuGlyIleIleArgValAlaAspGlyAsnLysGlyValPheThr 612  
 Db 30329 GATTTTGCACCTCGCAAGCATATTCGCCGTTGGCGGTAAACAAGAAAGGATCTTCACT 30270  
 QY 613 ArgAspArgLysProLysAlaAlaAlaHisSerLeuArgAlaArgTyrThrSerIleAsp 632  
 Db 30269 CCGAACCAGAAACGAAAGTGGCGGCTTTTCTGTGCAAAAAACGCTGACCTGGCATGAAC 30210

Search completed: March 22, 2005, 20:39:38  
 Job time : 966 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: March 22, 2005, 18:06:15 ; Search time 284 Seconds  
(without alignments)  
3652.815 Million cell updates/sec

Title: US-10-757-093-4

Perfect score: 3354  
Sequence: 1 MRFGLSLSLSLAPSLGTP.....RKPKAAHSLRMTSIDKN 634

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODS=frame.p2n.model -DB=xlp  
-Q/cgn2\_1/USPTO.epool\_p/US10757093/runat\_18032005\_164456\_27776/app\_query.fasta\_1.775  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=DELCL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10757093 @CNS\_1\_1.105 @runat\_18032005\_164456\_27776 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA.\*  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/prodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/prodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1721.5	51.3	3169	3	US-08-630-820-5
2	1721.5	51.3	3169	4	US-09-273-453-5
3	1714.5	51.1	32798	4	US-09-604-694B-1
4	1712.5	51.1	3035	2	US-08-723-624-18
5	1711	51.0	3824	2	US-08-723-624-19
6	1709	51.0	4652	4	US-09-893-525-36
7	1706	50.9	5390	4	US-09-893-525-41
8	1706	50.9	5418	4	US-09-893-525-38
9	1703	50.8	11978	4	US-09-792-568-8
10	1703	50.8	12438	4	US-09-792-568-9
11	1692	50.4	8012	3	US-09-182-117-1
12	1692	50.4	8012	4	US-09-434-039A-1

13	1692	50.4	8418	3	US-09-182-117-5	Sequence 5, Appl1
14	1692	50.4	8418	4	US-09-434-039A-5	Sequence 5, Appl1
15	1692	50.4	8798	3	US-09-182-117-4	Sequence 4, Appl1
16	1692	50.4	8798	4	US-09-434-039A-4	Sequence 4, Appl1
17	1673.5	49.9	2141	4	US-09-445-283C-27	Sequence 27, Appl1
18	1673.5	49.9	8076	4	US-09-532-806-2	Sequence 2, Appl1
19	1667.5	49.7	4947	4	US-09-118-276-21	Sequence 21, Appl1
20	1667.5	49.7	5642	1	US-08-118-772A-2	Sequence 2, Appl1
21	1667.5	49.7	5897	4	US-09-097-319A-26	Sequence 26, Appl1
22	1667.5	49.7	5897	4	US-09-643-971-26	Sequence 26, Appl1
23	1667.5	49.7	6898	3	US-09-097-319A-27	Sequence 27, Appl1
24	1667.5	49.7	6898	4	US-09-643-971-27	Sequence 27, Appl1
25	1667.5	49.7	9002	4	US-09-532-806-3	Sequence 3, Appl1
26	1667.5	49.7	9299	3	US-09-643-971-15	Sequence 15, Appl1
27	1667.5	49.7	9299	4	US-09-643-971-15	Sequence 15, Appl1
28	1667.5	49.7	9335	3	US-09-097-319A-19	Sequence 19, Appl1
29	1667.5	49.7	9335	4	US-09-643-971-19	Sequence 19, Appl1
30	1667.5	49.7	9408	3	US-09-097-319A-16	Sequence 16, Appl1
31	1667.5	49.7	9408	4	US-09-643-971-16	Sequence 16, Appl1
32	1667.5	49.7	10160	3	US-09-097-319A-8	Sequence 8, Appl1
33	1667.5	49.7	10160	4	US-09-643-971-8	Sequence 8, Appl1
34	1667.5	49.7	11784	3	US-09-097-319A-9	Sequence 9, Appl1
35	1667.5	49.7	11784	4	US-09-643-971-9	Sequence 9, Appl1
36	1667.5	49.7	11991	3	US-09-097-319A-10	Sequence 10, Appl1
37	1667.5	49.7	11991	4	US-09-643-971-10	Sequence 10, Appl1
38	1666	49.7	7742	2	US-08-882-704A-4	Sequence 4, Appl1
39	1666	49.7	7742	3	US-09-151-957-4	Sequence 4, Appl1
40	1664.5	49.6	4544	3	US-09-488-270A-1	Sequence 1, Appl1
41	1659	49.5	2438	6	5432081-1	Sequence 1, Appl1
42	1659	49.5	2438	6	5432081-1	Sequence 1, Appl1
43	1648.5	49.2	4284	1	US-08-525-507-14	Sequence 14, Appl1
44	1631.5	48.6	2633	1	US-08-452-267-2	Sequence 2, Appl1
45	1631.5	48.6	2633	3	US-09-123-644-2	Sequence 2, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-630-820-5  
; Sequence 5, Application US/08630820  
; Patent No. 6008023  
; GENERAL INFORMATION:  
; APPLICANT: OPPER, Martin  
; APPLICANT: BOSSLET, Klaus  
; APPLICANT: CZECH, Joerg  
; TITLE OF INVENTION: CYTOPLASMIC EXPRESSION OF ANTIBODIES,  
; TITLE OF INVENTION: ANTIBODY FRAGMENTS AND ANTIBODY FRAGMENT FUSION MOLECULES  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/630,820  
; FILING DATE: 10-APR-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE 19513676.4  
; FILING DATE: 11-APR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: GRANADOS, Patricia D.  
; REGISTRATION NUMBER: 33,683  
; REFERENCE/DOCKET NUMBER: 18748/306  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300

TELEFAX: (202) 672-5399  
 TELEEX: 904136  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3169 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: circular  
 MOLECULE TYPE: cDNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Enterobacteriaceae: Escherichia coli  
 STRAIN: PR4210  
 IMMEDIATE SOURCE:  
 CLONE: pTrc99 dicistr. Fab/E.c.-Beta-Gluc  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 3..641  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 666..3162  
 US-08-630-820-5

Alignment Scores:  
 Pred. No.: 3 02e-175 Length: 3169  
 Score: 1721.50 Matches: 339  
 Percent Similarity: 68.17% Conservative: 98  
 Best Local Similarity: 52.89% Mismatches: 179  
 Query Match: 51.33% Indels: 25  
 Gaps: 11

US-10-757-093-4 (1-634) x US-08-630-820-5 (1-3169)  
 QY 8 SerLeuSerLeuLalaProSerLeuGlyThrProAla-----AlaArg 23  
 Db 1233 AGGGTGTGATCGTGGCCCTCCAGACGCTTGGGACCCAGACTACACCTGCAAGTGAAT 1292  
 QY 24 HisPheProArgSerGluMetThrGlnHisGluGlnProLeu-----Ile 38  
 Db 1293 CACAGACCCGACGACACCCAGAGTGAGACAGAGATTGAGCTCGCTTCTGCTGCTCC 1352  
 QY 39 LysValArgProGlnArgThrSerSerArgLysLeuValAsnLeuAspGlyLeuTrpLys 58  
 Db 1353 ATGGTACGTCTGTGAAACCCCAACCCGCGAATCAAAAATCTCAGCGCTGTGGCA 1412  
 QY 59 PheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTrpThrAlaPro 75  
 Db 1413 TTCAGTCTGGATCGCGAAACTGTGGAAATGATCAG-----CGTTGTTGGAAAGCGCG 1466  
 QY 76 LeuProLysGlyLeuGluCysProValProAlaSerTrpAsnAspIlePheIleSerArg 95  
 Db 1467 TTTCAGAAAGCCGCGCAATTGCTGTGTCAGGCGATTTAACATCACTTCGCCGATGCA 1526  
 QY 96 GlnIleHisAspHisValGlyTrpValTrpValTrpGlnArgGluValIleValProLysGly 115  
 Db 1527 GAAATTCGTAATTATGCGGCAACGTCGTGATCAAGCGCAAGCTTTATACCGAAAGGT 1586  
 QY 116 TrpSerGlnGluArgTrpLeuValArgAlaGluSerAlaThrHisHisGlyArgIleTyr 135  
 Db 1587 TTGGGAGCGCGAGTATCGTGTGCTGCTTCGATCGGCTGACTCATTCACGCGCAAGGTGTG 1646  
 QY 136 ValAsnAsnArgLeuValAlaGluHisValGlyGlyTrpThrProPheGluValAsnVal 155  
 Db 1647 GTCAATATATCAGAAAGATGATGAGCATCAGGCGGCTATACGCCATTGTAAGCGGATGTC 1706  
 QY 156 ThrGluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyValAsnAsnGluLeu 175  
 Db 1707 ACCGCGATATATTATGCGGGAAGATGATACGATACCGTTGTGTGGAACAAGAACTG 1766  
 QY 176 ThrHisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysArgIle 195  
 Db 1767 AACTGGAGAGCTATTCGCGCGGGAATGATGATTAC---GACGAAACGCGCAAGAAAG 1823

QY 196 GlnThrTrpGlnHisAspPheTrpAsnTrpAlaGlyLeuAlaArgSerIleTrpLeuTyr 215  
 Db 1824 CAGTCTTACTTCAATAATTTCTTAACCTATGCGGGAATCCATCGACGCTATGCTTAC 1883  
 QY 216 SerValProGlnGlnHisIleGlnAspIleThrValIleThrAspVal-----AspGly 233  
 Db 1884 ACCAGCGCGAACACCTGGGTGACCGATATCACCGTGTGACCATATGTCGCGCAAGCTGT 1943  
 QY 234 AspAsnGlyLeuIleAsnTrpGluValGluValAlaAsnGlnThrThrGlyGlnIleGln 253  
 Db 1944 AACCAAGCGCTGTGTGACTCGGAGGTG---GTGGCAAT-----GTATGTCAAC 1991  
 QY 254 IleSerValIleAspGluLysAspGlyAlaIleValAlaIleValAlaSerGlyValaGlnGlyThr 273  
 Db 1992 GTTGAACCTCGTATGCGGATCAACAGTGTGCACTGCAACAGCAAGCACTAGCGGACT 2051  
 QY 274 ValThrIleProSerValLysLeuTrpGlnProGlyAlaAlaIleTyrLeuTrpGlnLeuGln 293  
 Db 2052 TTGCAAGTGTGAATCCGACCTGCGCAACCGGATGAAGTTATCTTATGAACTGTGC 2111  
 QY 294 ValAsnIleValGlySerSerGlyAspValValaAspThrTrpAsnLeuAlaThrGlyVal 313  
 Db 2112 GTC-----ACAGCCAAAGCCAGACAGATGTGATCTACCCGCTCGCGCATC 2165  
 QY 314 ArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysProPheTrpPheThr 333  
 Db 2166 CGGTACGTGCGATGAAAGGCGCAACGCTTCTGATTAACCAACAACCGTTCTTACTTACT 2225  
 QY 334 GlyPheGlyLysHisGluAspThrAlaValaArgGlyLysGlyHisAspProAlaTrpMet 353  
 Db 2226 GGCTTGTGTGTCATGAATCCGACTTACCTGCGCAAGATTCGATTAACGTGTGATGAG 2285  
 QY 354 ValHisAspPheGlnLeuMetLysTrpIleGlyValaAsnSerPheArgThrSerHisTyr 373  
 Db 2286 GTGCAGACCGACCATTAATGACGTGATTTGAGGCGCAACTCTTACCGTACCTGCAATAC 2345  
 QY 374 ProTyrAlaGluGluValMetAspPheAlaAspArgSerGlyIleValValIleAspGlu 393  
 Db 2346 CCTTACGCTGAAGATGCTGACCTGCGCAATGAACATGAGCATGCTGTGATTTGATA 2405  
 QY 394 ThrProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGluSerGlyAlaPro 412  
 Db 2406 ACTGCTGCTGTGGCTTTAACTCTCTTACGATGTTGCGAAGCGGCAACAAGCGCG 2465  
 QY 413 ---GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAlaHisLysGlnAla 431  
 Db 2466 AAGAACTGTACAGCAAGAGCGATCAACGGGGAATCTACGACGCACTTACAGGCG 2525  
 QY 432 IleArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMetTrpSerIleAla 451  
 Db 2526 ATTAAGACTGATAGCGGTGATCAAAAACCCCAAGGTGTGATGTGAGATTTGCG 2585  
 QY 452 AsnGluProAlaSerHisGluAspGlyAlaArgGluTyrPheGluProLeuThrAsnLeu 471  
 Db 2586 AACGAACCGGATACCCGTGCGCAAGGTGCAAGGAAATATTTGCGGCACCTGCGGAAGA 2645  
 QY 472 ThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyr 491  
 Db 2646 ACCGCTAACTGACCCGACGCGGTCCGATCACCTGCGTCAAAAGTATGTTTCGCAAGCT 2705  
 QY 492 GlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTrp 511  
 Db 2706 CACACGATATCATCAGCATCTCTTGAATGTGTGTGCTGTAACCGTTATTAACGATGG 2765  
 QY 512 TyrSerGlnThrGlyAspLeuGluGluAlaGluAlaIleLeuGluLysGluLeuHisGly 531  
 Db 2766 TATGTCCAAAGCGCGGATTTGAAACGCAAGAGTCTGGAAAAAACAATCTTGTGGCC 2825  
 QY 532 TrpGlnGluLysPheHisArgProIleValMetThrGluTyrGlyValaAspThrLeuAla 551  
 Db 2826 TGGCAGAGAAATGTCATCAGCGGATTAATCATCAACGAATACGCGCTGTGATACGTTAGCC 2885

QY 552 GlyLeuHisSerIleLeuGlyLeuProTrpSerGluGluPheGluValGlnMetLeuAsp 571  
 Db 2886 GGGCTGACCTCATGTATACCGACCATGTGAGTAAAGATATCATGTGATGGCTGGAT 2945  
 QY 572 MetTyrHisArgValPheAspArgIleGluSerMetAlaGlyGluHisValTTPAspPhe 591  
 Db 2946 ATGTATCACCGCGCTTTGATTCGCGCAGCGCCGCTCGTGAAACGATATGGAATTTTC 3005  
 QY 592 AlaAspPheGlnThrLeuGlyIleLeuArgValAspGlyValGlyValPhe 611  
 Db 3006 GCGGATTTTGCAGCTCGCAAGCATATTCGCGTGGCGTAACAAGAAAGGATTTTC 3065  
 QY 612 ThrArgAspArgLysProLysAlaAlaHisSerLeuArgAlaArgTPTThrSerIle 631  
 Db 3066 ACTCGGACCGCAACCGAAGTGGCGGCTTTTCTGCTGCAAAAACCGCTGATCGGATG 3125  
 QY 632 Asp 632  
 Db 3126 AAC 3128  
 RESULT 2  
 US-09-273-453-5  
 ; Sequence 5, Application US/09273453  
 ; Patent No. 6602688  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OPPER, Martin  
 ; BOSSLER, Klaus  
 ; CZECH, Joerg  
 ; TITLE OF INVENTION: CYTOPLASMIC EXPRESSION OF ANTIBODIES,  
 ; ANTIBODY FRAGMENTS AND ANTIBODY FRAGMENT FUSION MOLECULES  
 ; IN E. COLI  
 ; NUMBER OF SEQUENCES: 7  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 3000 K Street, N.W., Suite 500  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20007-5109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/273,453  
 ; FILING DATE: 22-Mar-1999  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/630,820  
 ; FILING DATE: <Unknown>  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: GRANADOS, Patricia D.  
 ; REGISTRATION NUMBER: 33,683  
 ; REFERENCE/DOCKET NUMBER: 18748/306  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 672-5300  
 ; TELEFAX: (202) 672-5399  
 ; TELEX: 904136  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3169 base pairs  
 ; TYPE: nucleic acid  
 ; MOLECULE TYPE: circular  
 ; TOPOLOGY: circular  
 ; STRANDEDNESS: double  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Enterobacteriaceae: Escherichia coli  
 ; STRAIN: pRAU210  
 ; IMMEDIATE SOURCE:  
 ; CLONE: pTrc99 dicistr. Fab/E.c.-Beta-Gluc  
 ; FEATURE:

NAME/KEY: CDS  
 LOCATION: 3..641  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 666..3162  
 SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
 US-09-273-453-5  
 Alignment Scores:  
 Pred. No.: 3.02e-175 Length: 3169  
 Score: 1721.50 Matches: 339  
 Percent Similarity: 68.17% Conservative: 98  
 Best Local Similarity: 52.89% Mismatches: 179  
 Query Match: 51.33% Indels: 25  
 DB: 4 Gaps: 11  
 US-10-757-093-4 (1-634) x US-09-273-453-5 (1-3169)  
 QY 8 SerLeuLeuSerLeuAlaAlaProSerLeuGlyThrProAla-----AlaArg 23  
 Db 1233 AGCGTGATACCGTCCGCTTCACAGCAGCTTGGCAACCAACCTACACAGCGAAT 1292  
 QY 24 HisPheProArgGlnGluMetThrGlnHisGluGlnProLeu-----Ile 38  
 Db 1293 CACAAGCCCGACGACACCAAGGTGACAAAGAGATTGAGCTCGTTCTGGTTCGTTCC 1352  
 QY 39 LysValArgProGlnArgThrSerArgGluLeuValAsnLeuAspGlyLeuTPLYs 58  
 Db 1353 ATGGTACGTCCTGTAGAAACCCCAACCCCGTAATCAAAAACCTGACGCGCTGGGCA 1412  
 QY 59 PheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTPTThrAlaPro 75  
 Db 1413 TTCAGCTGGATGCGCAAAACTGTGAATTGATGAC-----CGTTGGTGGAAAGCGG 1466  
 QY 76 LeuProLysGlyLeuGluCysProValProAlaSerTyrAsnAspIlePheIleSerArg 95  
 Db 1467 TTACAAGAAAGCGGGCAATTGCTGTCCAGCAGGAGTTTAAACATGATCAGTCCGATGCA 1526  
 QY 96 GluIleHisAspHisValGlyTTPValIYTYGlnArgGluValIleValProLysGly 115  
 Db 1527 GATATTGTAATTATTCGGGCAACGCTGTGTACACGGCAAGCTTTTATACGAAAGT 1586  
 QY 116 TrpSerGlnGluArgYrLeuValArgAlaGluSerAlaThrHisIleGlyArgIleTyr 135  
 Db 1587 TGGGACGGCCAGCGTATTCGCTGCGCTTCGATCGCGCTCATATTACGCAAGTGTGG 1646  
 QY 136 ValAsnAspArgLeuValAlaGluHisValGlyIYTYrThrProPheGluAlaAspVal 155  
 Db 1647 GTCAATATACAGAAAGTATGAGCATCAGGGCGGTATACCGCATTTGAAAGCGATGTC 1706  
 QY 156 ThrGluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyValAsnAspGluLeu 175  
 Db 1707 ACCCGCTATGTATTCGCCGGAAGTGAAGTATACCGTTGTGTGAACAAGACTG 1766  
 QY 176 ThrHisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysArgIle 195  
 Db 1767 AACTGGACAGCTATCCCGCGGGAATGTGATTAAC---GACGAAACCGCAAGAAAG 1823  
 QY 196 GlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaAspSerIleTTPLeuTyr 215  
 Db 1824 CAGCTTACTTCATTAATTTCTTTAATTATGCGGGATGATGCGAGCGTAATGCTTAC 1883  
 QY 216 SerValProGlnGlnHisIleGlnAspIleThrValIleThrAspVal-----AspGly 233  
 Db 1884 ACCACCGCGAACAACCTGGGTGTGACGATATCACCGTGTGACGATGTCCGCAAGCTGT 1943  
 QY 234 AspAsnGlyLeuIleAsnTyrGluValAlaAsnGluThrThrGlnGlnIleGln 253  
 Db 1944 AACCAACGCTGCTTACTGCGAGTG---GTGGCAAT-----GATATGTACAGC 1991  
 QY 254 IleSerValIleAspGluAspGlyAlaIleValAlaLysAlaSerGlyAlaGlnGlyThr 273  
 Db 1992 GTTGAAGCTGGTATGCGGATCAACAGTGTGGCAATGGAACAAGCACTACGCGGACT 2051

```

QY 274 ValThrIleProSerValIleuThrProGluValAlaIleuThrGluLeuGln 293
   :::::
Db 2052 TTGCAGAGTGTAAATCCGCACTCTGGCAACCGGGTGAAGGTTATCTTATGAACTGTGC 2111
QY 294 ValAsnIleValGlySerSerGlyAspValValAspThrTyrAsnIleAlaThrGlyVal 313
   2112 GTC-----ACAGCCAAAGCCAGACAGAGTGTATCTACCCCGCTTCGGGTGGCATC 2165
QY 314 ArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysProPheTyrPheThr 333
   2166 CGGTAGTGGCAGTGAAGGGCGAAGATTCCTGATTAAACCAAAACCGTTCTACTTAACT 2225
QY 334 GlyPheGlyLysHisGluAspThrAlaValArgGlyLysGlyHisAspProAlaTyrMet 353
   2226 GGCCTTGCTGTCATGAAGATCGCGCACTTACGTGGCAAGATTCGATACCTCTGATG 2285
QY 354 ValHisAspPheGlnLeuMetLysTrpIleGlyAlaAsnSerPheArgThrSerHisTyr 373
   2286 GTGCACGACCAACGCAATTATGAGATGAGATGGGGCCAACTCTACCGTACCTGGCATTAC 2345
QY 374 ProTyrAlaGluGluValMetAspPheHisAspArgAsnGlyIleValIleAspGlu 393
   2346 CTTACGCTGAAGATGCTCTGCACTGGCAGATGAACATGCGATCGTGATTTGATGA 2405
QY 394 ThrProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGluSerGlyAlaPro 412
   2406 ACTGCTGCTGCGGCTTTAACTCTCTTTAGGCATGTGTTTCGAAGGGGCAACAAACCG 2465
QY 413 ---GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAlaHisLysGlnAla 431
   2466 AAAGAACTGTCACAGAAAGCGCAAGCTGACCGGAAACTGACGAAGGCACTTACACGGCG 2525
QY 432 IleArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMetTrpSerIleAla 451
   2526 ATTAAGAGCTGATAGCGCGTGAACAACCAACCGCTGTGTGATGTGAATATTTGCC 2585
QY 452 AsnGluProAlaSerHisGluAspGlyAlaArgGlyLysThrPheGluProLeuThrAsnLeu 471
   2586 AACCAACCGGATACCCGCTCGCAAGGTGACGCGGAATATTTCCGCGCATCGCGGAAGCA 2645
QY 472 ThrArgGlnLeuAspProThrArgProIleThrPheHisAsnValGlyThrAlaThrTyr 491
   2646 ACGGTAACTCAACCGGACGCGCTCGCATCACTGCGCTCAATGTAATGTTTGGCAGCGT 2705
QY 492 GlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTrp 511
   2706 CACACCGATCCCATACGCGATCTCTTGATGTGCTGTGCTGAACCGTTATTAACGATGG 2765
QY 512 TyrSerGlnThrGlyAspLeuGluGluAlaGluAlaLeuGluLysGluLeuHisGly 531
   2766 TATGTCAAAGCGCGCGATTGGAAACGCGCAGAGAGTACTGGAAAAAACAACCTTGGCC 2825
QY 532 TrpGlnGluLysPheHisAspProIleValMetThrGlyTyrGlyAlaAspThrIleAla 551
   2826 TGGCAGAGAAATCGCATCAGCGCATTAATCAACCGAATACCGCGTGTGATACGTTAGCC 2885
QY 552 GlyLeuHisSerIleLeuGlyLeuProTrpSerGluGlnPheGlnValGlnMetLeuAsp 571
   2886 GGGCTCACTCAATCAATGACCGCATGTGAGTGAAGATATCAAGTGTGATGGCTGGAT 2945
QY 572 MetTyrHisArgValPheAspArgIleGluSerMetAlaGlyGlnHisValIleAspPhe 591
   2946 AGTATATCAACCGGCTCTTGATCCGCTACACCGCGCTGTGTGTAACAGGTATAGGAATTC 3005
QY 592 AlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnLysLysGlyValPhe 611
   3006 GCGGATTTTGCACCTTCGCAAGGCAATATGCGCGTTGGCGGTAAACAAGAAAGATCTTC 3065
QY 612 ThrArgAspArgLysProLysAlaAlaHisSerLeuArgAlaArgTrpThrSerIle 631
   3066 ACTCGGACCGCAAAACCGAAGTGGGGCTTTTCTGTGCAAAAACGCTGGACGTGGCATG 3125

```

```

QY 632 Asp 632
Db 3126 AAC 3128

RESULT 3
US-09-604-694B-1/c
; Sequence 1, Application US/09604694B
; Patent No. 6579522
; GENERAL INFORMATION:
; APPLICANT: BROUGH, DOUGLAS E
; APPLICANT: KING, C R
; APPLICANT: KOVESDI, IMRE
; APPLICANT: SCHAIBLE, JASPER J
; TITLE OR INVENTION: REPLICATION DEFICIENT ADENOVIRAL TNF VECTOR
; FILE REFERENCE: 202028
; CURRENT APPLICATION NUMBER: US/09/604,694B
; NUMBER OF FILING DATE: 2003-01-10
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 32798
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
US-09-604-694B-1

Alignment Scores:
Pred. No.: 8,196-173
Score: 1714.00
Percent Similarity: 70.50%
Best Local Similarity: 55.17%
Query Match: 51.10%
DB: 4
Gaps: 9

US-10-757-093-4 (1-634) x US-09-604-694B-1 (1-32798)

QY 40 ValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpLysPhe 59
Db 31982 GTCCCTCTCTGTAAGAAACCCCAACCGGTGAATCAAAAACTGACGCGCTGTGGCATTC 31923
QY 60 AlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTrpThrAlaProLeu 76
Db 31922 AGTCGTGATCCGAAACTGTGGAATTGATGACG-----CGTTGTGGGAAGCGCGT 31869
QY 77 ProLysGlyLeuGluCysProValProAlaSerTyrAsnAspIlePheIleSerArgGlu 96
Db 31868 CAAAGAACCCGGGCAATGCTGTGCGCAGGCGATTTAACGATCAGTTCGCCGATGCAGAT 31809
QY 97 IleHisAspHisValGlyTrpValTyrTyrGlnArgGluValIleValProLysGlyTrp 116
Db 31808 ATTGCTAATTATGGGGGCAAGCTGTGATACGCGCAAGTCTTTATACCGAAAGTTGG 31749
QY 117 SerGlnGluArgTyrLeuValArgAlaGluSerAlaThrHisHisGlyArgIleTyrVal 136
Db 31748 GCAGGCCAGCGCTATCGCGCTGCGCTTTCGATCGCGTCACTCATTAACGCAAAAGTGGGTC 31689
QY 137 AsnAspArgLeuValAlaGluHisValGlyTyrThrPheProPheGluAlaAspValThr 156
Db 31688 AATAATACGAAGATGATGAGACATCAGGGCGGCTATAGCGCATTTGAAGCGGATGTCACG 31629
QY 157 GluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyValAsnAsnGluLeuThr 176
Db 31628 CCGTATGTATTATGGCGGGAAGATGATATACCGCTTTGTGTGAACAACGAAGTGAAC 31569
QY 177 HisGlnThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysArgIleGln 196
Db 31568 TGGCAGACTATCCCGCGGGAATGGTGAATTAAC--GACCAAAACGCGCAAGAAAGCGCG 31512
QY 197 ThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIleTrpLeuTyrSer 216
Db 31511 TCTTACTTCATATGATTTCTTAACTATGCGGAAATCATCGCAGCGTATGCTCTTACACC 31452

```

QY 217 ValProGlnGlnHisIleGlnAspIleThrValValThrAspVal-----AspGlyAsp 234  
 DB 31451 AGCCGGAACACCGTGGTGGATATCACGCTGTACCGATGTCGCGCAAGCTGCAAC 31392  
 QY 235 AsnGlyLeuIleAsnTyrgluValGluValAlaAsnGlnThrGlyGlnIleGlnIle 254  
 DB 31391 CAGCGCTGTGTGACCTGGCAGGTG---GTGGCCAT-----GGTATGTCACCGGT 31344  
 QY 255 SerValIleAspGluAspGlyAlaIleValAlaLysAlaSerGlyValGlnGlyThrVal 274  
 DB 31343 GAACCTCGATGCGGATCAACAGGTGTGCACTGACGACAGGCACTACGCGGACTTGG 31284  
 QY 275 ThrIleProSerValIleLeuTrpGlnProGlyValAlaAlaTyrgluLeuGlnVal 294  
 DB 31283 CAAGTGTGATTCGCAACCTTGGCAACCGGTGAAGTATTCTATAGACCTGTGCTC 31224  
 QY 295 AsnIleValGlySerSerGlyAspValValAspThrTyrgluLeuAlaThrGlyValArg 314  
 DB 31223 -----ACAGCCAAAGCCAGACAGGTGATATATACCCGCTGCGCGCATCCGG 31170  
 QY 315 ThrValIleValAlaIleGlySerGlnPheLeuIleAsnGlyLysProPheThrGly 334  
 DB 31169 TCAGTGGCAGTGAAGGCGCAACGCTTGATTACCAACCGTTCATTACTTGGC 31110  
 QY 335 PheGlyLysHisGluAspThrAlaValArgGlyLysGlyHisAspProAlaTyrgluVal 354  
 DB 31109 TTTGGTGTGATTAATGATGCGGACTTGGCGCAAGGATTCATACCGTGTGATGCTG 31050  
 QY 355 HisAspPheGlnLeuMetLysTrpIleGlyAlaAsnSerPheArgThrSerHisTyrglu 374  
 DB 31049 CACGACCAAGCATTAATGATGCTGATTTGGGGCCACTCTTACCGTACCTGCTACCT 30990  
 QY 375 TyrAlaGluGluValMetAspPheAlaAspArgAsnGlyLysValValIleAspGluThr 394  
 DB 30989 TACGCTGAAGAGATGCTCACTGGCAAGATGAACGATGCTGTGATGATGAAGAACT 30930  
 QY 395 ProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGluSerGlyAlaPro--- 412  
 DB 30929 GCTGCTGTGGCTTAACTCTCTTAAAGCATGCTTCAACCGGCAACAGCCGAA 30870  
 QY 413 GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAlaHisGlyGlnAlaIle 432  
 DB 30869 GAACCTGACGCGAAGGAGGACGCAACGCGGAAACTCAGCAACGCGCACTTACGCGAT 30810  
 QY 433 ArgGluLeuIleAlaIleArgAspLysAsnHisAlaSerValValMetTrpSerIleAlaAsn 452  
 DB 30809 AAAGACTGATAGCGGTGCAAAAACCAACCAAGCGTGTGATGTGATATTTGCCAAC 30750  
 QY 453 GluProAlaSerHisGluAspGlyAlaArgGluTyrgluPheGluProLeuThrAsnLeuThr 472  
 DB 30749 GAACCGGATACCGCTCGCGCAAGGTGCAAGGAAATATTTCCGCGCACTGCGCGAAGCA 30690  
 QY 473 ArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyrgln 492  
 DB 30689 CGTAACTCGACCGCGAAGCGCTCGATCACTCGCTCATATATGTTTCGCAAGCGTCAC 30630  
 QY 493 LeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrgluTyrgluTyrglu 512  
 DB 30629 ACCGATACCATCAGCATCTTGTGATGTGCTGTGCTGCAACCGCTTATTACGATGTAT 30570  
 QY 513 SerGlnThrGlyAspLeuGluGluValGluAlaIleGlnGluGluGluGluGluGluGlu 532  
 DB 30569 GTCCAAAGCGGCGATTTGAAACGCGAAGAGTACTGAAAAAGAACTTTCGCGCTGG 30510  
 QY 533 GlnGluLysPheHisArgProIleValMetThrGlyTyrglyValIleAspThrLeuAlaGly 552  
 DB 30509 CAGGAAGAACTGATCAGCGGATTAATCAACCGAATACCGCGTGTGATAGCCGG 30450  
 QY 553 LeuHisSerIleLeuGlyLeuProTrpSerGluGluPheGlnValGlnMetLeuAspMet 572  
 DB 30449 CTGCACTCATATGACCGCAACATGTGAGAGTGAAGATATCATGTGCATGCGTGTGATATG 30390  
 QY 573 TyrHisArgValPheAspArgIleGluSerMetAlaGlyGluHisValIleTrpAsnPheAla 592

DB 30389 TATCACCGCGCTTTGATGCGTCAGCGCGCTGCTGATGGAACAGGATGAAATTTCCGC 30330  
 QY 593 AspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnLysLysGlyValPheThr 612  
 DB 30329 GATTTTGCACCTTCGCAAGGCAATTTGGCGGTGGCAAGAAAGGATCTTCACT 30270  
 QY 613 ArgAspArgLysProLysAlaAlaHisSerLeuArgAlaArgTrpThrSerIleAsp 632  
 DB 30269 CGGACCGCAACCGAAGTCGGCGCTTTTCGCTGCAAAAACGCTGACATGCAAGAC 30210  
 RESULT 4  
 US-08-723-624-18  
 ; Sequence 18, Application US/08723624  
 ; Patent No. 5861277  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rose, Alan B.  
 ; APPLICANT: Laet, Robert L.  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ENHANCING  
 ; TITLE OF INVENTION: THE EXPRESSION OF GENES IN PLANTS  
 ; NUMBER OF SEQUENCES: 21  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: ARNOLD, WHITE & DURKEE  
 ; STREET: P.O. Box 4433  
 ; CITY: Houston  
 ; STATE: TX  
 ; COUNTRY: USA  
 ; ZIP: 77210  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/723,624  
 ; FILING DATE: Concurrently Herewith  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Parker, David L.  
 ; REGISTRATION NUMBER: 32,165  
 ; REFERENCE/DOCKET NUMBER: BTIP:002  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 512/418-3000  
 ; TELEFAX: 512/474-7577  
 ; INFORMATION FOR SEQ. ID NO: 18:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3035 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-08-723-624-18  
 Alignment Scores:  
 Pred. No.: 2,64e-174 Length: 3035  
 Score: 1712.50 Matches: 333  
 Percent Similarity: 69.23% Conservative: 99  
 Best Local Similarity: 53.37% Mismatches: 169  
 Query Match: 51.06% Indels: 23  
 DB: 2 Gaps: 10  
 US-10-757-093-4 (1-634) x US-08-723-624-18 (1-3035)  
 QY 16 SerLeuGlyThrProAlaAlaArgHisPheProArgAsnGluMetThrGlnHisGluGln 35  
 DB 1176 GCGGTGCGACGACGCGCTAGAGGATCCCGGCTGTCACTCC----- 1220  
 QY 36 ProLeuIleLysValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGly 55  
 DB 1221 -----CTTCTGTACGTCTCTGTGAACCCCAACCGTGAATACTCAACGCGC 1274  
 QY 56 LeuTrpLysPheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTrp 72  
 DB 1275 CTGTGGCATTTCACTGTGATGCGCAAACTGTGAATGATGAG-----CGTGGTGG 1328



QY 73 ThrAlaProLeuProIysGlyLeuGlyCysProValProAlaSerTyraAsnIlePhe 92  
 Db 1329 GAAGGCGCTTACAAAGAACCGGCAATTCCTGCTCCAGCGCTTTAAAGATCACTGTC 1388  
 QY 93 ILeSerArgLulIeHisAspHisValGlyTyrValTyrTyrGlnArgLulValIleVal 112  
 Db 1389 GCCGATGCAGATATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTAT 1448  
 QY 113 ProIysGlyTyrSerGlnGlnArgTyrLeuValArgAlaGluSerAlaThrHisGly 132  
 Db 1449 CGAAAGGTGGGCGACCGCATATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1508  
 QY 133 ArgIleTyrValAsnAsnArgLeuValAlaGluHisValGlyTyrThrProPheGlu 152  
 Db 1509 AAAGTGTGGTCAATATACAGAAAGTATGAGACATCAAGGCGCATATACCGCATTTGAA 1568  
 QY 153 AlaSerValThrGluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyValAsn 172  
 Db 1569 GCCGATGTACAGCGCATATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTAT 1628  
 QY 173 AsnGluLeuThrHisGlyThrThrIleProProGlyLysIleThrThrGlyAsnAlaThrGly 192  
 Db 1629 AACGAATGAATCTGCGAAGCTATCCGCGGAGATGATGATTAAC--GACGAAACGGC 1685  
 QY 193 LysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuValaArgSerIle 212  
 Db 1686 AAGAAAGACAGCTTACTTCCATGATTTCTTAATTCGCGAATCCATGCGACAGTA 1745  
 QY 213 TrpLeuTyrSerValProGlnGlnHisIleGlnAspIleThrValValThrAspVal--- 231  
 Db 1746 ATGCTCTACACCCCGCGAACACCTGCGTGCAGATATCACCTGTGTGACGCTGCGG 1805  
 QY 232 ---AspGlyAspAsnGlyLeuIleAsnTyrGluValGluValAlaAsnGlnThrGly 250  
 Db 1806 CAAGATGTATACACCGCTCTGTGACGTGCGAGTC---GTGCGAAT-----GGT 1853  
 QY 251 GlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAlaLysAlaSerGlyAla 270  
 Db 1854 GATGTACAGGTTGAATCGCGGTATGCGGATCAACAGGTGTTGCAATGGAACAAGCACT 1913  
 QY 271 GlnGlyThrValThrIleProSerValLysLeuTyrGlnProGlyAlaAlaTyrLeuTyr 290  
 Db 1914 AGCGGACATTTGGCAATGCGAATCCGCACTCTGCGCAACCGGTAAGATTATCTCTAT 1973  
 QY 291 GlnLeuGlnValAsnIleValGlySerSerGlyAspValValAspThrTyrAsnLeuAla 310  
 Db 1974 GAACCTGCGTC-----ACAGCCAAAGCCAGACAGATGATATCTAACCGCTTCGC 2027  
 QY 311 ThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysProPhe 330  
 Db 2028 GTGCGCATCCGGTCACTGAGAGTGAAGGGGAACAGTCTCTGATTAACCAACCAACGTTG 2087  
 QY 331 TyrPheThrGlyPheGlyLysHisGluAspThrAlaValArgGlyLysGlyHisAspPro 350  
 Db 2088 TACTTACCTGCGCTTGTGCTCATGAAGATGCGCGATGTGCGCAAGATTCGATTAAC 2147  
 QY 351 AlaTyrMetValHisAspPheGlnLeuMetLysTyrIleGlyValaAsnSerPheArgThr 370  
 Db 2148 GTGCTATGTGACACACACGATTAATGAGATGAGATTTGGGCGCAATCTTACCGTACC 2207  
 QY 371 SerHisTyrProTyrAlaGluGluValaMetAspPheAlaAspArgAsnGlyIleVala 390  
 Db 2208 TCGCATTAACCTTAACCTGAAGATGCTGCACTGGGCAAGTAAACATGCGCATCGTGGTG 2267  
 QY 391 IleAspGluThrProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGluSer 409  
 Db 2268 ATTGATGAACACTGCTCTGTGCGCTTAACTCTCTTAGGCACTGCGTTTCGAAGCGGCG 2327  
 QY 410 GlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAlaHis 428  
 Db 2328 AACACGCCGAAGAACTGTACAGCGAAGGCAAGTCAACGGGAAACCTACGACGACGAC 2387

QY 429 LysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMetTyr 448  
 Db 2388 TTACAGCGCATTAAGAGCTGATAGCGCTGACCAAAAAACCAACCGATGATGTGG 2447  
 QY 449 SerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGluTyrPheGluProLeu 468  
 Db 2448 AGTATTTGCCAAGAACCGGATACCCGTCGCAAGGTGACGGGAATATTTCCGCGCACGTG 2507  
 QY 469 ThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThr 488  
 Db 2508 GCGGAAGCAACCGGTAACTCAACCGACGCTCCGATCACTGCTCAATGAATGCTTC 2567  
 QY 489 AlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyr 508  
 Db 2568 TCGAGCGCTCACACCGATACCATACCGCATCTCTTGAATGTCTGTGCTGCAACCGTTAT 2627  
 QY 509 PheGlyTyrTyrSerGlnThrGlyAspLeuGluValaGluAlaAlaLeuGluLysGlu 528  
 Db 2628 TACGATGGTATGTCCAAAGCGCGCATTTGGAAACGGCAGAAAGTACTGGAAAAAGAA 2687  
 QY 529 LeuHisGlyTyrGlnGluLysPheHisArgProIleValaMetThrGluTyrGlyAlaAsp 548  
 Db 2688 CTTCGCGCTGCGCAGAGAACTGCATCAACCGCATATCATCAACGAATACCGCGTGAAT 2747  
 QY 549 ThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyrSerGluGluPheGlnValGln 568  
 Db 2748 ACGTATGCGCGGCTGACATCAATGACACCGCATGTGAGTGAAGATATCATGATGCA 2807  
 QY 569 MetLeuAspMetTyrHisArgValPheAspArgIleGluSerMetAlaGlyGluHisVal 588  
 Db 2808 TCGCTGATATGATATCACCGCGCTTTGATCCGCTACGCGCGCTGCTGCTGAACAGTAT 2867  
 QY 589 TrpAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValaAspGlyAsnLysLys 608  
 Db 2868 TGAATTTGCGCATTTTCGACCTCGCAAGCATATTCGCGGTGCGGTAACAAGAA 2927  
 QY 609 GlyValPheThrArgAspArgLysProLysAlaAlaHisSerLeuArgAlaArgTyr 628  
 Db 2928 GGGATTTCACTGCGCACCGCAACCGAAGTGGCGGCTTTCTGCTGCAAAAAACGCTGG 2987  
 QY 629 ThrSerIleAsp 632  
 Db 2988 ACTGCGATGAAC 2999

RESULT 5  
 US-08-723-624-19  
 ; Sequence 19, Application US/08723624  
 ; Patent No. 5861277  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rose, Alan B.  
 ; APPLICANT: Laet, Robert L.  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ENHANCING  
 ; TITLE OF INVENTION: THE EXPRESSION OF GENES IN PLANTS  
 ; NUMBER OF SEQUENCES: 21  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: ARNOLD, WHITE & DUREKE  
 ; STREET: P.O. Box 4433  
 ; CITY: Houston  
 ; STATE: TX  
 ; COUNTRY: USA  
 ; ZIP: 77210  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/723,624  
 ; FILING DATE: Concurrently Herewith  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Parker, David L.  
 ; REGISTRATION NUMBER: 32,165

REFERENCE/DOCKET NUMBER: BTIP:002  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 512/418-3000  
 TELEFAX: 512/474-7577  
 INFORMATION FOR SEQ ID NO: 19:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3824 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-723-624-19

Alignment Scores:  
 Pred. No.: 5,56e-174 Length: 3824  
 Score: 1711.00 Matches: 330  
 Percent Similarity: 70.50% Conservative: 93  
 Best Local Similarity: 55.00% Mismatches: 161  
 Query Match: 51.01% Indels: 16  
 Gaps: 9

US-10-757-093-4 (1-634) x US-08-723-624-19 (1-3824)

```

QY 40 ValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpLysPhe
DB 2016 TTAGCTGCTGTAGAAACCCCAACCCGGAATCAAAAACCTGCGCGCTGTGGGCAATTC
QY 60 AlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTrpThrAlaProLeu
DB 2076 AGCTCGATCGCAAAAACCTGGAATGATGAC-----CGTTGGTGGGAAAGCGCTTA
QY 77 ProLysGlyLeuGlyCysProValProAlaSerTyrAsnAspIlePheIleSerArgGlu
DB 2130 CAAGAAACCGCGCAATTCCTGTCGCCAGCACTTTTAAACATCACTTCCCGCATCGAT
QY 97 IleHisAspHisValGlyTrpValTyrTyrGlnArgGluValIleValProLysGlyTrp
DB 2190 ATTGTAATATATGCGGGCAACGCTGGTATCAGGCGCAAGCTTTATACCGAAAGGTTCG
QY 117 SerGlnGluArgTyrLeuValArgAlaGluSerAlaThrHisHisGlyArgIleTyrVal
DB 2250 GCGAGCGAGGTCGTCGTCGCTTCGATGCGTCACTCATTAACGCAAAAGTGGTTC
QY 137 AsnAspArgLeuValAlaGluHisValGlyGlyTyrThrProPheGluAlaAspValThr
DB 2310 AATTAATCAGAAATGATGAGCATCAGGCGGCTATTCGCAATTTGAAGCCGATGTCACG
QY 157 GluLeuValAlaProGlyGlyLysPheArgLeuThrIleGlyValAsnAspGluLeuThr
DB 2370 CGGTATGTTATGCGCGGAAAGGTATGATACCGTTGTTGTGAAACGAACGATGAC
QY 177 HisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysValGlnIleGln
DB 2430 TGGCAGACTATCCCGCGGGAATGCGATTAAC---GACGAAAAACGCAAAAGAAACGAC
QY 197 ThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIleTrpLeuTyrSer
DB 2487 TCTTACTTCATATTTCTTTAATCTATGCGGAATTCATGACGCAAGTAAGCTCTTACAC
QY 217 ValProGlnGlnHisIleGlnAspIleThrValValThrAspVal-----AspGlyAsp
DB 2547 ACCCGCAACACCGGGGCGAGCATATCACCGTGTGACGATCGCGCAAGACTGTAAC
QY 235 AsnGlyLeuIleAsnTyrGluValGluValAlaAsnGlnThrThrGlyGlnIleGlnIle
DB 2607 CACGCGCTGCTGACTGCGCAGGTG---GTGGCCAAAT-----GGTGAATGTCAGCGTT
QY 255 SerValIleAspGluAspGlyAlaIleValAlaLysAlaSerGlyAlaGlnGlyThrVal
DB 2655 GAATGGTGTATCGGATCAACAGGTGTGCACTGCAACAGGCACTACGGGACTTTC
QY 275 ThrIleProSerValLysLeuTrpGlnProGlyAlaAlaIleTyrLeuTyrGlnLeuGlnVal
DB 2715 CAAGTGTGTATCGGACCTCTGCGCAACCGGGTGAAGATTATCTCTATGAAGTGTGCTC
  
```

```

QY 295 AsnIleValGlySerSerGlyAspValValAspThrTyrAsnLeuAlaThrGlyValArg
DB 2775 -----ACAGCCAAAGCCAGACAGATGATATCTACCCGCTTCCGCGCATCCGG
QY 315 ThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysAspProPheTyrPheThrGly
DB 2829 TCAGTGGCAGTGAAGGCGCAAGATTCCGTATTAACCAAAACGTTCTACTTATTCGCG
QY 335 PheGlyLysHisGluAspThrAlaValArgGlyLysGlyHisAspProAlaTyrMetVal
DB 2889 TTGGTGTATCAATGAATGGAGCTGCGCGCAAGATTCAATACGTCGATGATGTCG
QY 355 HisAspPheGlnLeuMetLysTrpIleGlyAlaAsnSerPheArgThrSerHisTyrPro
DB 2949 CACGACCAACCATTAATGACATGATGGGCGCAACTCTCAACCTGATCTGCACTTACCT
QY 375 TyrAlaGluGluValMetAspPheAlaAspArgAsnGlyIleValValIleAspGluThr
DB 3009 TACGCTGAAGATAGCTCGACCTGGCAGATGAACATGCAATCGTGTGATGTAACACT
QY 395 ProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGluSerGlyAlaPro---
DB 3069 GCTGCTGTGCGCTTTAACTCTCTTAAGCATTTGTTGCAACCGGCAACAGCCGAAA
QY 413 GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluValHisLysGlnAlaIle
DB 3129 GAACCTGACAGCAAGAGGAGTCAAGGGGAAACTCAGCAACGCACTTACAGCGATT
QY 433 ArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMetTrpSerIleAlaAsn
DB 3189 AAGAGCTGATACGGGCTGACAAACCAACCAAGGTGTGATGTGGAATATTCGCAAC
QY 453 GluProAlaSerHisGluAspGlyAlaArgGlyTyrPheGluProLeuThrAsnLeuThr
DB 3249 GAACCGGATACCCGTCGCAAGGTGCAAGGAATATTTCCGCGACATCGCGGAACAGC
QY 473 ArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyrGln
DB 3309 CGTAACCTGACCAACCGAGCGCTCGATCACTGCTCATATTAATGTTCTCCAGCCCTAC
QY 493 LeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTrpTyr
DB 3369 ACCGATACCATCAGCATCTCTTGAATGTCTGTGCTGCAACCGTTATTCAGATGTAT
QY 513 SerGlnThrGlyAspLeuGluValAlaGluAlaIleGluLysGluLeuHisGlyTrp
DB 3429 GTCCAAAGCGCGATTTGAAACGCGAGAGAGTACTGGAAGAAACCTTCGCGCTCG
QY 533 GlnGluLysPheHisArgProIleValMetThrGluTyrGlyValAlaAspThrLeuAlaGly
DB 3489 CAGAGAAATCGATACAGCGATTAATCATCAACGAATACCGGCTGTGATGATGACCGGG
QY 553 LeuHisSerIleLeuGlyLeuProTrpSerGluGluPheGlnValGlnMetLeuAspMet
DB 3549 CTGCACTCAATGATACACCGCATGTGAGTGAAGATATAGTGTGATGCTGCAATATG
QY 573 TyrHisArgValPheAspArgIleGluSerMetAlaGlyGluHisValIleTrpAsnPheAla
DB 3609 TATCACCGCGCTTGTATCGCTGACGCGCTGTGTGTGAACAGATATGAATTTTCGCC
QY 593 AspPheGlnThrAsnLeuGlyIleIleArgValAlaAspGlyValAsnLysGlyValAlaPheThr
DB 3669 GATTTTGGCAGCTCGCAAGCATATTTGGCGGTGGCGGTAAAGAAAGGATCTTCACT
QY 613 ArgAspArgLysProLysAlaAlaHisSerLeuArgAlaArgTrpThrSerIleAsp
DB 3729 CGGACCGCAACCGAAGTGGCGCTTTTCTCTCTGCAAAACGCTGATGCGCATGAAC
  
```

RESULT 6  
 US-09-893-525-36  
 ; Sequence 36, Application US/09893525  
 ; Patent No. 6753167



Db 2949 CGTTATTACGATGTATGTCCAAAGCGCGATTTGGAAACGCGAGAAAGTACTGAA 3008  
 QY 527 LysGluLeuH1s61YTrpGlnGluYpRheH1sArgProIleValMetThrGluYrGly 546  
 Db 3009 AAGAACTTCTGCGCTGGCAGGAAACTGTGATGAGCGATATCATACCGAAATACGGC 3068  
 QY 547 AlaSerThrLeuH1s61YLeuH1sSerIleLeuGlyLeuProTrpSerGluGluPheGln 566  
 Db 3069 GCGGATACGTTACCGGGGCGCATCATATACACGCAACGTGAGATGAAGATACAG 3128  
 QY 567 ValGluMetLeuH1s61YrH1sArgValPheAspArgIleGluSerMetAlaGlyGlu 586  
 Db 3129 TGTGATGCGTGGATATGATACCGCGCTTTGATGCGCTGACGCGCTGCGGTGA 3188  
 QY 587 HisValTrpAspRheH1s61YrH1sArgValPheAspArgIleGluSerMetAlaGlyGlu 606  
 Db 3189 CAGGATACGATATCCCGCATTTTGGACCTCGCAAGCGATATGCGCTGCGGTGAAC 3248  
 QY 607 LysLysGlyValPheThrArgAspArgLysPheLysAlaIleH1sSerLeuArgAla 626  
 Db 3249 AAGAAAGGATCTTCACTCGCGAACCGAAACCGAAGTCCGCGCTTTCTGCTGCAAAA 3308  
 QY 627 ArgTrpThrSerIleAsp 632  
 Db 3309 CGCTGACTGCGATGAAAC 3326  
 RESULT 7  
 US-09-893-525-41  
 / Sequence 41, Application US/09893525  
 / Patent No. 6753167  
 / GENERAL INFORMATION:  
 / APPLICANT: Moloney, Maurice M.  
 / APPLICANT: Van Rooijen, Gije  
 / TITLE OF INVENTION: Preparation of Heterologous Proteins on Oil Bodies  
 / FILE REFERENCE: 9369-172  
 / CURRENT APPLICATION NUMBER: US/09/893,525  
 / PRIOR FILING DATE: 2001-06-29  
 / PRIOR APPLICATION NUMBER: US 09/210,843  
 / PRIOR FILING DATE: 1998-12-15  
 / PRIOR APPLICATION NUMBER: US 08/846,021  
 / PRIOR FILING DATE: 1997-04-25  
 / PRIOR APPLICATION NUMBER: US 08/366,783  
 / PRIOR FILING DATE: 1994-12-30  
 / PRIOR APPLICATION NUMBER: US 08/142,418  
 / PRIOR FILING DATE: 1993-11-16  
 / PRIOR APPLICATION NUMBER: US 07/659,835  
 / PRIOR FILING DATE: 1991-02-22  
 / NUMBER OF SEQ ID NOS: 42  
 / SOFTWARE: PatentIn version 3.1  
 / SEQ ID NO 41  
 / LENGTH: 5390  
 / TYPE: DNA  
 / ORGANISM: Artificial Sequence  
 / FEATURE:  
 / OTHER INFORMATION: phase-caleo-GUS-phas  
 / NAME/KEY: CDS  
 / LOCATION: (1548)..(4097)  
 / OTHER INFORMATION:  
 / US-09-893-525-41  
 Alignment Scores:  
 Pred. No.: 3 34e-173 Length: 5390  
 Score: 1706.00 Matches: 329  
 Percent Similarity: 70.50% Conservative: 94  
 Best Local Similarity: 54.83% Mismatches: 161  
 Query Match: 50.86% Indels: 16  
 Gaps: 9  
 US-10-757-093-4 (1-634) x US-09-893-525-41 (1-5390)  
 QY 40 ValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpLysPhe 59  
 Db 2292 TTACGTCCTGTAGAAACCCCAACCCGCTGAATAAATACTCAACGCGCTGTGGCATTC 2251

QY 60 AlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTrpThrAlaProLeu 76  
 Db 2352 AGCTGATCGCGAAAACTGTGAATTGATCAG-----CGTTGCTGGAAACGGCTTA 2405  
 QY 77 ProLysGlyLeuGluCysProValProAlaSerTyraAsnAspIlePheIleSerArgGlu 96  
 Db 2406 CAAGAAAGCCGGCAATTGCTGTGCCAGGCAAGTTTAAAGATCAGATCTGCCGATGCAAT 2465  
 QY 97 IleHisAspHisValGlyTrpValTyTrpGlnArgGluValIleValProLysGlyTrp 116  
 Db 2466 ATTCTGAATTAATTCGGGCAACGCTGTGTACAGCGGAAGTCTTATACGGAAGGTGG 2525  
 QY 117 SerGlnGluArgTyTrpValAlaArgAlaGluSerAlaThrHisHisArgIleTyVal 136  
 Db 2526 GCAGCCAGCGATATGCTGCTGCTTTCATGCGGTCACTCATTCAGCAAGGTGGTGC 2585  
 QY 137 AsnAsnArgLeuValAlaGluHisValGlyTyTrpThrProPheGluAlaAspValThr 156  
 Db 2586 AATTAATCAGGAAGTATGAGCATCAGGCGGCTATACGCATTTGAAGCCGATGTCAG 2645  
 QY 157 GluLeuValAlaProGlyGluYpRheArgLeuThrIleGlyValAsnAsnGluLeuThr 176  
 Db 2646 CGCTATGTTATTCGCGGAAAGTGTACGATCACCGCTTGTGTGAACAACGAATGAAAC 2705  
 QY 177 HisGluThrIleProProGlyLysIleThrGlyAsnAlaThrGlyLysArgIleGln 196  
 Db 2706 TGGCAGACTATCCCGCGGAAATGTGATTAAC--GACGAAACCGCAAGAAACGAG 2762  
 QY 197 ThrTyGlnHisAspPheTyraAsnTyraGlyLeuAlaArgSerIleTrpLeuTySer 216  
 Db 2763 TCTTACTTCATGATTTCTTTAACTATGCGGAATCCATCGCAGCGTAATGCTTACAC 2822  
 QY 217 ValProGlnGlnHisIleGlnAspIleThrValIleThrAspVal-----AspGlyAsp 234  
 Db 2823 ACCGCCAACAACCTGGGTGAGCATATCACCGGTGTACGATCCGATCCGCAAGCTGAAC 2882  
 QY 235 AsnGlyLeuIleAsnTyrgluValGluValAlaAsnGlnThrThrGlnIleGlnIle 254  
 Db 2883 CAGCGCTGTGTACCTGGCAGGTG--GTGGCAAT-----GATATGTCAAGCTT 2930  
 QY 255 SerValIleAspGluAspGlyAlaIleValAlaIleValAlaSerGlyAlaGlnGlyThrVal 274  
 Db 2931 GAACGCGATGATCGGATCAACAGCGTGTGAACGCAAGCACTGAGCGGACCTTGG 2990  
 QY 275 ThrIleProSerValLysLeuTrpGlnProGlyAlaIleTyTrpLeuTyGlnLeuGlnVal 294  
 Db 2991 CAAGTGTGAATCCGACCTCTGGCAACCGGGTGAAGTATCTATGAAACTGTGCGTC 3050  
 QY 295 AsnIleValGlySerSerGlyAspValValAspThrTyraAsnLeuAlaThrGlyValArg 314  
 Db 3051 -----ACAGCCAAAGCCAGACAGAGTGTGATATCTACCGCTTCCGCTGGCATCCG 3104  
 QY 315 ThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysProPheTyTrpThrGly 334  
 Db 3105 TCAGTGCAGTGAAGGCCAAGCAAGTTCCTGATTAACCAAAACCTTCACTTATCTGCG 3164  
 QY 335 PheGlyLysHisGluAspThrAlaValArgGlyLysGlyHisAspProAlaTyMetVal 354  
 Db 3165 TTGTGTCGTCAATGATGCGGACCTTACGTCGCAAGATTCGATTAACGTGCTGATGCTG 3224  
 QY 355 HisAspPheGlnLeuMetLysTrpIleGlyAlaAsnSerPheArgThrSerHisTyPro 374  
 Db 3225 CACGACACGACATTAATGATCGGATTTGGGCCCAACTCTTACCGTATCCGTCGATTAACCT 3284  
 QY 375 TyraLysGluValMetAspPheAlaAspArgAsnGlyIleValValIleAspGluThr 394  
 Db 3285 TTACGTCGAAGATGATCGATCTGCGCAATGAACGATGATCTGTGATTTGAAGAACT 3344  
 QY 395 ProAlaValGlyLeuAsnIleAlaLeu--MetGlyValSerGluSerGlyAlaPro--- 412  
 Db 3345 GCTGCTGTGCGGCTTTTCCCTCTTTAGAGCATTTGTCGAAGCGGCAACGAGCCGAA 3404

```

QY 413 GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAlaHisIleSglnAlaIle 432
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3405 GAACGTATACAGCAGAGGACGACGACGAGGAACTCAGCAAGCGCATTTACAGCGGATT 3464
QY 433 ArgGluLeuIleAlaArgAspLysAsnHisAlaSerValIleMetTrpSerIleAlaAsn 452
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3465 AAAGAGCTGATAGCGCGTGAACAAACCAACCCAGCGGTGATGTGAGTATTCGCCAAC 3524
QY 453 GluProAlaSerHisSglnAspGluValIleArgGluTyrPheGluProLeuThrAsnLeuThr 472
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3525 GAACCGGATACCGCTCGCAGAGGTGACCGGAATATTTCCGCGCATTCGCGGAAGCAAG 3584
QY 473 ArgGluLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyrGln 492
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3585 CGTAACTGACCCGACCGCGCTCGCATCCTGCTCAATGTATGTTCTTCGCCAGCTCAC 3644
QY 493 LeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTyr 512
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3645 ACCGATACCATCAGCGATCTTTGATGTGCTGTGCTGCTGACCGTTATTACGATGTAT 3704
QY 513 SerGlnThrGlyAspLeuGluGluValIleGluAlaIleGluLeuHisSglnTyr 532
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3705 GTCCAAAGGGCGCATTTGAAACGCGAGAGAGTACTGGAAAAAGAACTTCGGCCTGG 3764
QY 533 GlnGluLysPheHisArgProIleValMetThrGluTyrGlyValIleAspThrLeuAlaGly 552
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3765 CAGGAGAACTGATCAGCCGATATCATCACCAGTAACCGCGTGAATACGTTAGCCGG 3824
QY 553 LeuHisSerIleLeuGlyLeuProTyrSerGluGluPheGlnValGlnMetLeuAspMet 572
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3825 CTGCACATCAATGATACCCGACATGTGAGAGTGAAGATACATGTCATGCGTGAATAG 3884
QY 573 TyrHisArgValPheAspArgIleGluSerMetAlaGlyGlnHisValIleThrAsnPheAla 592
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3885 TATCACCGGCTCTTATGCGCGACCGCGCTGCTGCTGCAACAGATAGAAATTTCCGC 3944
QY 593 AspPheGlnThrAsnLeuGlyIleIleArgValIleAspGlyAsnLysGlyValIlePheThr 612
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3945 GATTTTCCGACCTCGCAGAGCATATTCGCGTTGGCGGTACCAAGAGGATCTTCACT 4004
QY 613 ArgAspArgLysProLysAlaIleAlaHisSerLeuArgIleArgTyrThrSerIleAsp 632
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 4005 CGCGACCGGAAACCGAAGTCGCGCTTTCTGCTCAAAAACGCTGACATGACATGAC 4064

```

```

NAME/KEY: CDS
LOCATION: (1555)..(1908)
OTHER INFORMATION:
NAME/KEY: CDS
LOCATION: (2149)..(4125)
OTHER INFORMATION:
US-09-893-525-38

Alignment Scores:
Pred. No.: 3,36e-173
Score: 1706.00
Percent Similarity: 70.50%
Best Local Similarity: 54.83%
Query Match: 50.86%
DB: 4 Gaps: 9

US-10-757-093-4 (1-634) x US-09-893-525-38 (1-5418)
QY 40 ValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTyrLysPhe 59
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2320 TTACGTCCTGTGAAACCCCAACCCGTAATCAAAAACCTGACGCGCTGTGGGCAATTC 2379
QY 60 AlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTyrThrAlaProLeu 76
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2380 AGCTCGATCGCGAAACCTGTGAATGATCAG-----CGTTGTGGGAAAGCGGCTTA 2433
QY 77 ProLysGlyLeuGluLysProValProAlaSerTyrAsnAspIlePheIleSerArgLys 96
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2434 CAGAAAGCCGGGCAATTGCTGTGCGACAGTTTAAAGATCACTTCCCGCATGCAGAT 2493
QY 97 IleHisAspHisValGlyTyrValTyrTyrGlnArgGluValIleValProLysGlyTyr 116
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2494 ATTGCAATTAATGCGCGGCAACGTCGTGATCAGCGGAAAGTCTTATACGAAAGTTGG 2553
QY 117 SerGlnGluArgTyrLeuValIleArgAlaGluSerAlaThrHisHisGlyArgIleTyrVal 136
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2554 GCAGCGCACCGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2613
QY 137 AsnAspArgLeuValAlaGlnHisValGlyGlyTyrThrProPheGluAlaAspValThr 156
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2614 AATATCAGAGAAATGATGAGATCAGGCGGCTATACCCATTGAAACCGATGTCAG 2673
QY 157 GluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyValAsnAsnGluLeuThr 176
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2674 CCGTATGTTATTCGCGGAAAGTGAAGTATACCATCAGCTTGTGTGAACAAGAACTGAAC 2733
QY 177 HisGluThrIleProProGlyLysIleThrTyrGlyAsnAlaThrGlyLysArgIleGln 196
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2734 TGGCAGACTATCCCGCGCGGAATGGATTAAC-----GACGAAACCGCAAGAAAGCAG 2790
QY 197 ThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIleThrPheLysSer 216
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2791 TCTTATTCATGATTTCTTAACTATGCGGAATCATCGACGCTAATGCTTACAC 2850
QY 217 ValProGlnGlnHisIleGlnAspIleThrValIleThrAspVal-----AspGlyAsp 234
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2851 AGCGCGAACAACCTGGGTGACATATCACCGGTGACGATGTCGCGCAAGCTGTATAC 2910
QY 235 AsnGlyLeuIleAsnTyrGluValGluValAlaAsnGlnThrThrGlyGlnIleGlnIle 254
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2911 CACGCGCTGTGTGACCGCAGGAGG--GTGGCCAA-----GCTATGTACAGCGTT 2958
QY 255 SerValIleAspGluAspGlyAlaIleValAlaLysAlaSerGlyValaGlnGlyThrVal 274
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2959 GAACGTCGTGATGCGGATCAACAGGTGTGCAACTGCAAGGACACTGCGGACTTTG 3018
QY 275 ThrIleProSerValLysLeuTyrGlnProGlyAlaAlaTyrLeuTyrGlnLeuGlnVal 294
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3019 CAAGTGTGAATCCGACCTCGCAACCGGGTGAAGTATCTATGAACTGTGCGTC 3078
QY 295 AsnIleValGlySerSerGlyAspValValaAspThrTyrAsnLeuAlaThrGlyValaArg 314
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3079 -----ACAGCCAAAGCCAGACAGAGTGTGATATCTACCCGCTTCGCGCATCCGG 3132

```

```

QY 315 ThrValIyValAlaGlySerGlnPheLeuIleAsnGlyLysProPheThyGly 334
Db 3133 TCAGTGGACGTAGAGGCGCAAGCTTCCTATTAACACAAACCGTTCTTACGCGC 3192
QY 335 PheGlyLysHisGlyLysPheThyAlaValArgGlyLysGlyHisPheProAlaTyMetVal 354
Db 3193 TTTGGTGGTCATGAGATGGGACCTTAACGTGGCAAGAGATTGATTAACGTGATGGTG 3252
QY 355 HisAspPheGlnLeuMetLysTrpIleGlyAlaAsnSerPheArgThySerHisTyPro 374
Db 3253 CACGACCCAGCATTAATGACATGGATGGGCGCAACCTCAACCTGACCTGCAATTCCT 3312
QY 375 TyrAlaGlyLysValMetAspPheAlaAspArgAsnGlyIleValValIleAspGlyThr 394
Db 3313 TACGCTGAAGATAGTCTGACATGGGACATGAACATGGCATCGTGGATGATTAAGAACT 3372
QY 395 ProAlaValGlyLysAsnIleAlaLeu---MetGlyValSerGlyLysPro--- 412
Db 3373 GCTGCTGTGGGCTTTTGGCTCTCTTAAAGCATGTTTGAAGCGGCAACAGCGGAAA 3432
QY 413 GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAlaHisLysGlnAlaIle 432
Db 3433 GAACGTACAGCGAAGAGCGACGTCACCGGGAACCTCAGCAAGCGCACTTACAGCGCAT 3492
QY 433 ArgGlyLeuIleAlaArgAspLysAsnHisAlaSerValIleMetTrpSerIleAlaAsn 452
Db 3493 AAGAGCTGATACCGGCTGACAAACACCAACCGGTGGATGGAGATTTGGCCAC 3552
QY 453 GluProAlaSerHisGlyLysPheGlyAlaArgGlyLysPheGlyLysProLeuThrAsnLeuThr 472
Db 3553 GAACCGGATACCGCTCGCAAGAGTGACCGGGAATATTTCGCGCACTGGGGAAGCAACG 3612
QY 473 ArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyGln 492
Db 3613 CGTAACTGACACCGGACCGCTCGATCACCCTGCTCATATGTTCTTCGACGCTCAC 3672
QY 493 LeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyPheGlyTrp 512
Db 3673 ACCGATACCTACCGGATCTTGTGATGCTGCTGCTGCAACCGTTATTAACGATGGAT 3732
QY 513 SerGlnThrGlyAspLeuGluGluAlaGluAlaIleAsnGlyLysGlyLysGlyLys 532
Db 3733 GTCGAAGCGGCGATTTGGAAACGCAAGAAAGTACTGAAAGAACTTCGCGCTGG 3792
QY 533 GlnGluLysPheHisArgProIleValMetThrGlyLysGlyLysAlaSerThrLeuAlaGly 552
Db 3793 CAGGAGAACTGCATCAGCCGATTAATCATCACCAGAAATACGCGGTGATAGCCGGG 3852
QY 553 LeuHisSerIleLeuGlyLeuProTrpSerGluGluPheGlnValGlnMetLeuAspMet 572
Db 3853 CTCGACTCATATGACACCGACATGAGTGAAGAGATATCATGTCATGGCTGGATATG 3912
QY 573 TyrHisArgValPheAspArgIleGluSerMetAlaGlyGluHisValIleTrpAsnPheAla 592
Db 3913 TATCAACCGGCTCTTATGATCGGTCAGCGCGCTGCTGATGACAGATATGATTTGCC 3972
QY 593 AspPheGlnThrAsnLeuGlyIleIleArgValAspGlyLysLysGlyValIlePheThr 612
Db 3973 GATTTCGCGCTCTGCAAGGCGATATTCGCGTTCGCGTAAAGAAAGGATCTTCACT 4032
QY 613 ArgAspArgLysProLysAlaAlaHisSerLeuAlaArgIleTrpThrSerIleAsp 632
Db 4033 CGGACCGCAACCGAAGTGGCGGCTTTCTCTGCAAAAACGCTGGACGATGAAAC 4092

```

```

; TITLE OF INVENTION: Other Organisms
; FILE REFERENCE: 2051US
; CURRENT APPLICATION NUMBER: US/09/792,568
; CURRENT FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FaSTSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 11978
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: plasmid pBG363 - binary vector with the nrg2 gene
US-09-792-568-8

Alignment Scores:
Pred. No.: 2,52e-172 Length: 11978
Score: 1703.00 Matches: 336
Percent Similarity: 68.60% Conservative: 101
Best Local Similarity: 52.75% Mismatches: 178
Query Match: 50.78% Indels: 22
DB: Gaps: 11

US-10-757-093-4 (1-634) x US-09-792-568-8 (1-11978)

QY 3 PheLeuThrGlyLysSerLeuLeuSerLeuAlaAlaProSerLeuGlyThrProAlaAla 22
Db 3844 TATGTTACGTCGGGTACATTTATTAACAGTGCAGATTAATCTTACCCCATTTT 3785
QY 23 ArgHisPheProArgAsnGluMetThrGlnHisGluGlnProLeuIleLysValArgPro 42
Db 3784 TACCATCTT-----GATCGAGATTAATACCGGTAAATTAATTCAGTA----- 3740
QY 43 GlnArgThrSerSerArgGlyLysLeuValAsnLeuAspGlyLysLeuTrpLysPheAlaLeu--- 61
Db 3739 ---GAAACCCCAACCGGGAATCAAAAACCTGACGCGCTGTGGCATTGCTGGAT 3683
QY 62 -----AlaSerGlyLeuAsnAspThrAlaGlnProThrAlaProLeuProLysGly 79
Db 3682 CGGAAACCTGTGAATGATGACG-----CGTTGGTGGAAAGCGGTTTCAAGAAAC 3629
QY 80 LeuGluCysProValProAlaSerTyAsnAspIlePheIleSerArgGluIleHisAsp 99
Db 3628 CGGCAATGCTGTGCGACGAGCAGTTTAACGATGACGTGCGATGACATATTCGTAAT 3569
QY 100 HisValGlyTrpValTyTyTyGlnArgGluValIleValProLysGlyTrpSerGlnGlu 119
Db 3568 TATGCGGCAACCTGTGATACCGCGAAGTCTTAATACCGAAGGTGGGCGACGCG 3509
QY 120 ArgTyLeuValArgAlaGlyLysSerAlaThrHisGlyValArgIleTyValAsnAsnArg 139
Db 3508 CGTATGCTGCTGCTTTCGATGCGGTCATCTCATTTACCGCAAACTGTGGTCAATATCAG 3449
QY 140 LeuValAlaGlyLysValGlyGlyTyThrProPheGluAlaAspValThrGluLeuVal 159
Db 3448 GAGTGTAGGAGCATCAGGCGGCTATACGCAATTTGAAGCCGATGACAGCGGTATGTT 3389
QY 160 AlaProGlyLysLysPheArgLeuThrIleGlyValAsnAsnGluLeuThrHisGlyThr 179
Db 3388 ATTCGCGGGAAGGTGTACATCACCGCTTGTGTAAACAACCAACGTAAGCTGACAGACT 3329
QY 180 IleProGlyLysLysIleThrThrGlyAsnAlaThrGlyLysArgIleGlnThrTyGln 199
Db 3328 ATCCGCGGGAATGTGATTAAC---GACGAAACGCGCAAGAAAGCAAGCTTACTTC 3272
QY 200 HisAspPheTyAsnTyAlaGlyLysLeuAlaArgSerIleTrpLeuTySerValProGln 219
Db 3271 CATGATTTCTTAATATGCGCGAATCAATCGCAGCGTATGCTTAACACCAAGCGGAAAC 3212
QY 220 GlnHisIleGlnAspIleThrValIleThrAspVal-----AspGlyAspAsnGlyLeu 237
Db 3211 ACCTGGGTGACATATCAACCGGTGACGCAATGTCGCGCAACAGCTGAACCGCGCTCT 3152
QY 238 IleAsnTyGluValGluValAlaAsnGlnThrThrGlyGlnIleGlnIleSerValIle 257

```

```
Db 3151 GTTGACTGGCAGGTG---GTGGCAAT-----GGTATGTCCAGCGTTGAACCTGCGT 3104
Qy 258 AspgluaspGlyAlaIleValAlaIleValAsnSerGlyAlaGlnGlyThrValThrIlePro 277
Db 3103 GATCGGATCAACAGCGGTGGTGGCAACTGGACACAGCACTACCGGGGAGCTTGGCAAGTGGTG 3044
Qy 278 SerAllylsLeuTPGlnProGlyAlaAlaIleValIleValIleValIleValIleVal 297
Db 3043 AATCGGACCTCTGGCAACCGGGTGAAGTTATCTTCTATGAACCTGTGCGTC-----ACA 2990
Qy 298 GlySerSerGlyAspValAlaIleValIleValIleValIleValIleValIleValIleVal 317
Db 2989 GCCAAAGCCAGACAGATGATATCTACCCGCTGCGCGGATCCGGTCAAGTGCA 2930
Qy 318 ValAlaIleValIleValIleValIleValIleValIleValIleValIleValIleVal 337
Db 2929 GTGAAGGCGCAACGCTTCTGATTAACCAACCGCTTCTTACTTGGCTTGTGCTG 2870
Qy 338 HisGluAspThrAlaValAlaIleValIleValIleValIleValIleValIleValIleVal 357
Db 2869 CATGAAGATGGCGACTTGGCGTGGCAAGATGATGATACGCTGATGCTGACGACAC 2810
Qy 358 GlnLeuMetIleValIleValIleValIleValIleValIleValIleValIleValIleVal 377
Db 2809 GCATTAATGACTGGATTGGGGCCCACTCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2750
Qy 378 GluValMetAspPheAlaAspArgAsnGlyIleValAlaIleValIleValIleValIleVal 397
Db 2749 GAGATGCTCACTGGCGACATGAACATGATGCTGATGATGATGATGATGATGATGATGATG 2690
Qy 398 GlyLeuAsnIleAlaIleValIleValIleValIleValIleValIleValIleValIleVal 415
Db 2689 GCGCTTTAAGCTCTTGAAGCATTTGTTGGAAGCGGCAACAGCGCAAGCAAGCAAGCAAGCT 2630
Qy 416 ThrProAspAlaIleValIleValIleValIleValIleValIleValIleValIleValIleVal 435
Db 2629 AGCGAAGAGGAGGAGTCAACCGGGAACCTGACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCT 2570
Qy 436 IleAlaIleValIleValIleValIleValIleValIleValIleValIleValIleValIleVal 455
Db 2569 ATAGCGCGTCAACAAACCAACCAAGCGTGTGATGTGATGTGATGTGATGTGATGTGATGTG 2510
Qy 456 SerHisGluAspGlyAlaIleValIleValIleValIleValIleValIleValIleValIleVal 475
Db 2509 ACCGCTCCGCAAGGTCACCGGAATTTTCGCGCACTGGCGGCAACCAAGCGCTAAACTC 2450
Qy 476 AspProThrArgProIleThrPheAlaAsnValGlyThrAlaIleValIleValIleValIleVal 495
Db 2449 GACCCGACCGGTCGATCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2390
Qy 496 IleSerAspLeuPheAspValSerGlyIleValIleValIleValIleValIleValIleValIleVal 515
Db 2389 ATCAGGATCTCTTGTATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2330
Qy 516 GlyAspLeuGlnGluAlaIleValIleValIleValIleValIleValIleValIleValIleVal 535
Db 2329 GCGGATTTGGAAACGCAAGAAAGTATCGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2270
Qy 536 PheHisArgProIleValIleValIleValIleValIleValIleValIleValIleValIleVal 555
Db 2269 CTGCATCAGCCGATTAATCAACCAAGTACCGCGTGAATCTTACCGCGGCTGCACTCA 2210
Qy 556 IleLeuGlyLeuProIleValIleValIleValIleValIleValIleValIleValIleValIleVal 575
Db 2209 AGTACACCGACATGGAGTGAAGATCACTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 2150
Qy 576 ValIlePheAspArgIleGluSerMetAlaGlyIleValIleValIleValIleValIleValIleVal 595
Db 2149 GCTCTTTGATCGGACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2090
Qy 596 ThrAsnLeuGlyIleIleValIleValIleValIleValIleValIleValIleValIleValIleVal 615
```

```
Db 2089 ACCTCCGACGCAATATTGGCGCTTGGCGGTACACAGAAAGGATCTTCACTGCGACCGC 2030
Qy 616 LysProIleValAlaIleValIleValIleValIleValIleValIleValIleValIleVal 632
Db 2029 AAACCGAAGTGGCGGCTTTCTGCTGCAAAAAAGCTGATCGGCAATGAC 1979

RESULT 10
US-09-792-568-9/c
; Sequence 9, Application US/09792568
; Patent No. 6696621
; GENERAL INFORMATION:
; APPLICANT: Klotz, Andreas
; APPLICANT: Davis, Keith
; TITLE OF INVENTION: No. 6696621el Selectable Marker in Plants and
; FILE REFERENCE: 2051US
; CURRENT APPLICATION NUMBER: US/09/792,568
; CURRENT FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 12438
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Plasmid pG361
US-09-792-568-9

Alignment Scores:
Pred. No.: 2,68e-172 Length: 12438
Score: 1703.00 Matches: 336
Percent Similarity: 68.60% Conservative: 101
Best Local Similarity: 52.75% Mismatches: 178
Query Match: 50.78% Gaps: 22
Db: 4 Indels: 11

US-10-757-093-4 (1-634) x US-09-792-568-9 (1-12438)
Qy 3 PheLeuThrGlyLeuSerLeuSerLeuAlaIleProSerLeuGlyThrProAlaIle 22
Db 3844 TATGTTCAGTCGCGTACATTTATTAATTAACAGTGCACATTAATTAATTAATTAATTAATTA 3785
Qy 23 ArgHisPheProArgAsnGluMetThrGlnIleGlnIleProLeuIleValIleValIleValIleVal 42
Db 3784 TACCATCTT-----GATCGATGATTAACCGTGTAAATTAATTAATTAATTAATTAATTAATTA 3740
Qy 43 GlnArgThrSerSerArgGluLeuValIleValIleValIleValIleValIleValIleValIleVal 61
Db 3739 ---GAAACCCCAACCGGGAATCAAAAACTGACACCGCTGTGGCATTCAGTCTGGAT 3683
Qy 62 -----AlaSerGlyLeuAsnAspThrAlaGlnProTPhrAlaProLeuProIleGly 79
Db 3682 CGCGAAACCTGGAATTAATGATCAG-----CGTTGGTGGAAACCGCTTACAAAGAAC 3629
Qy 80 LeuGluCysProValProAlaSerIleValIleValIleValIleValIleValIleValIleValIleVal 99
Db 3628 CGGCAATTCGTCGCGCAGGAGTTTAACGATCAGTTCCGCGATGACATATTCGATAT 3569
Qy 100 HisValGlyTPValIleValIleValIleValIleValIleValIleValIleValIleValIleVal 119
Db 3568 TATCGGCGCAACGCTGATCAAGCGGAGTCTTATCCGAAAGGTTGGCGAGGCGCAG 3509
Qy 120 ArgTyrLeuValAlaGluSerAlaThrHisIleGlyArgIleTyrValIleValIleValIleValIleVal 139
Db 3508 CGTATCGTCGCGCTTTCGATCGCGTCACTATTAACGCAAAAGTGGCTCAATTAATCAG 3449
Qy 140 LeuValAlaGluIleValIleValIleValIleValIleValIleValIleValIleValIleValIleVal 159
Db 3448 GAACTGATGAGCATCAGGCGGCTATACCGCAATTTGAACCGCATTCACGCGCTATGTT 3389
Qy 160 AlaProIleGlyIleValIleValIleValIleValIleValIleValIleValIleValIleValIleVal 179
Db 3388 ATTCCCGGAAAGTATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 3329
```



```

QY 180 ILEPROGILYLYSERTHRGLYSALATHRGILYSARGILEGINTHRYGILN 199
DB ATCCGCCCGGAGATGATGTTAC---GACGAAACGCGAGAAAGCATCTTACTTC 3272
QY 200 HISASPHETRYASNTYLAAGLYLEUALARGSERILETRPLEUTYRSEVALPROGIN 219
DB 3271 CATGATTTCTTAACTATGCGCGAATCCATCCAGGTAATGCTTCAACACGCGGAAAC 3212
QY 220 GLNHISILEGLINASPILETHRVALVALTHIRASPVAL-----ASPGLYASPAENGILYLEU 237
DB 3211 ACCTGGGTGACGATATCAACCGGTGACGATCGCGCAACATCTGAACCGCGGTCT 3152
QY 238 ILEASNTYRGLVALGLVALAIAENGINTHTRGLYGLINLEGINLIESERVALILE 257
DB 3151 GTTGACTGGCAGGTG---GTGGCCAAAT-----GGTATGATCAACCGTTGAACCGCT 3104
QY 258 ASPGLYASPGLYALALEVALALYVALASERGLYALAGINLYTHRVALTHIRILEPRO 277
DB 3103 GATGCGGATCAACAGGTGGTTGCACTGGACAGACGACTAGCGGACTTTGCAAGTGCTG 3044
QY 278 SERVALLYLEUTRPGINPROGILYALALATYRLEUTYRGLINLEUGLINVALASAILLEVAL 297
DB 3043 AATCCGACCTCTGGCAACCGGGTGAAGGTTATCTGATGAACTGTGCGCTC-----ACA 2990
QY 298 GLYSERSEGLYASPVALVALASPTHRYRASNLEUALATHRGILYVALARGTHRVALYS 317
DB 2989 GCCAAAAGCAGACAGAGTGTGATATCTACCCGCTCGCTCGCATCCGCTAGTGCGCA 2930
QY 318 VALALAGLYSERGLINPHELEULAEANGLYLYSPROPHETRYRPHETHRGLYPHEGLIYLS 337
DB 2929 GTGAAGGGGGAACAGTCTCTGATTAACCAACCGCTTACTTACTGCTTGGTGTGCT 2870
QY 338 HISGLIAPRTHRALVALARGGLYLYSGLYNHSAPPROALATYRMEVALHISAPRPH 357
DB 2869 CATGAAGATGCGGACTTGGCTGGCAAGATTCGATACCTGCTGATGTCGACGACCCAC 2810
QY 358 GLNLEUMETLYSTRILEGLYALAEANSERPHETHRSETHIRYRPROTYRVALAGILN 377
DB 2809 GCATTATATGACTGTGATGGGCAACTCTTACCGTACCTCGCATTAACCTTAAAGCTGAA 2750
QY 378 GLIUALMETASPHENALASPARANGLYLEVALIILEASPGILUTHRPROALVAL 397
DB 2749 GAGATCTCGACTGGGCAATGAACATGCGATCTGTGATTTGATAAAGCTGCTGCTC 2690
QY 398 GLYLEUASNTILEALALEU---METGLYVALSERGLISERGLYALAPRO---GLINTHRPHE 415
DB 2689 GGCTTTAACTCTCTTAAAGCATTTGTTGCAAGCGGCAACAGCCGAAAGAACTGTAC 2630
QY 416 THRPROASPIALAEASAPRYSTRHNGILUALHISILEGLINLALILEARGILULEU 435
DB 2629 ACGGAAAGAGGACTCAACGGGCAACTCGCAAGCCGCACTTACAGCGGATTAAGAGCTG 2570
QY 436 ILEALAEARGASPIYASNTISALASERVALVALMETRPSERTILEALAEANGIUPROALA 455
DB 2569 ATAGCGCGTGACAAAACCAACCAACCGTGTATGTGAGTATTCGCAACGAAACCGGAT 2510
QY 456 SERHISGLIAPSPILYALARGGLYLYRPHENGLUPROLEUTHRASNLEUTHRARGILULEU 475
DB 2509 ACCCGTCCCAAGGTGACAGCGGAAATTTTCGCCCACTGGCGGAAACCAACCGCTAAATC 2450
QY 476 ASPPRTHRARGPROILETHRPHENALAEANVALGLYTHRALATHRYRGLINLEUAPARG 495
DB 2449 GACCCGACCGCTCGATCAACCTGCGTCAATGTAATGTTCTGCAACCGCTCACACCGATACC 2390
QY 496 ILESERASPLEUPHEASPVASERCYHISLEASNTYRPHENGLYTRPYRSESGINTHR 515
DB 2389 ATCAGGAGATCTCTTATGATGCTGCTGCAACCGTTATTAACGATGATATGTCCAAAGC 2330
QY 516 GLIAPSPLEUGLILUALAGLUALAELEUGLILYSGILULEUHIISGLYTRPGINLILYLS 535
DB 2329 GGGGATTTGGAAACGCGACAGAGAGTACTGGAAAAAGAACTTCTGCGCTGGCAGAGAAA 2270

```

```

QY 536 PHEHISARGPROILEVALMETRTHRGILYLYALASPTHREUALAGLYLEUHSISER 555
DB 2269 CTGCATCAGCCGATTTATCATCAACGGAATACCGGTGATGATCGTTACCGCGCTGCACTCA 2210
QY 556 ILELEUGILYLEUPROTRPSERGLIUPHEGLINVALGLINMETLEUASPNECTYRHSARG 575
DB 2209 ATGTACACGACATGATGGAGATGAAGATGATCAAGTGTGATGCTGATATGATATCAACCGC 2150
QY 576 VALPHEASPARGLILEGLISERMELALAGLYLHISVALITRASNPHENALASAPRPHEN 595
DB 2149 GTCTTTGATTCGCTCGACCGCGTGTGGTGAACAGGTATGAAATTTCCCGGATTTTGGC 2090
QY 596 THRASNLEUGILYILELEARGVALAEPGLYASNTLYSGLYVALPHETHRARGAPARG 615
DB 2089 ACCTCCCAAGCATATATGGCGCTTGGCGGTGAACAAAGAAAGATCTTCACTCCGCAACCGC 2030
QY 616 LYSPPROLYALALAHISISERLEUARGALARGYTRPHTSERILEASP 632
DB 2029 AAACCGAAGTCGCGGCTTTTCTGCTGCAAAAAACGCTGACATGCAATGAAC 1979

RESULT 11
US-09-182-117-1
; Sequence 1, Application US/09182117
; Patent No. 6204436
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Transgenic Plant
; NUMBER OF SEQUENCES: 27
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/182,117
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ. ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8012 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-09-182-117-1

Alignment Scores:
Pred. No.: 2,056-171 Length: 8012
Score: 1692.00 Matches: 337
Percent Similarity: 68.32% Conservative: 95
Best Local Similarity: 53.32% Mismatches: 180
Query Match: 50.45% Indels: 20
DB: 3 Gaps: 11

US-10-757-093-4 (1-634) x US-09-182-117-1 (1-8012)
QY 10 LEUSERLEUALALAPROSERLEUGLYTHRPROIALALARGHISPHETROARGANGILU 29
DB 3601 CTATCTTCGCAAGACCTTCTCTATATGAAGATTCATTTGATGAGAGACAG 3660
QY 30 METTHGLNHISGLINPROLEUULEYVALARGPROGLINARGHISERASRARGILU 49
DB 3661 CTGACAAAGTGAATCTAGACATCTTCATGCTCCCTGTGAAACCCCAACCGCTGAA 3720
QY 50 LEUVALASNTLEUASPGILYLEUTRPLYSRPHENALALEU-----ALASERGLYLEUASN 66
DB 3721 ATCAAAAAACCTGACGCGCTGTGGCATTCAGCTGATCGCAAAACGTGAAATTTGAT 3780
QY 67 ASPTHRALAGINPROTRPHTRALAPROLEUPROLYSGILYLEUGILUCYSPROVALPROALA 86
DB 3781 CAG-----CGTTGTGGGAAAGCGCGTTTACAAAGAACCGGCAATTCCTGTGCGCAGGC 3834

```

```

QY      87 SerTyrAsnAspIlePheIleSerArgIleuIleHisAspHisValGlyTrrValTyrTyr 106
DB      3835 AATTTTAAGATCAGTGTCCGATGCGATATTCGAATTATTCGGGGCAACCTGTGAT 3894
QY      107 GlnArgGluValIleValProlysglyTrrSerGlnGluArgTyrLeuValArgAlaGlu 126
DB      3895 CAGCCGGAAGTCTTATACCGAAGGTGGCAGCGCAGCTATCGTGTGCGTTCCGAT 3954
QY      127 SerAlaThrHisHisGlyValIleTyrValAsnAspArgLeuValAlaGluHisValGly 146
DB      3955 GCGGTACTCTATTCACCGAAGGTGGCAGCTATTCGAATTATTCGGGGCAACCTGTGAT 4014
QY      147 GlyTyrThrProPheGluAlaAspValThrGluLeuValAlaProGlyGluLysPheArg 166
DB      4015 GGCTATACCGCATTTGAAGCCGATGTACGCCGTATGTTATTCGCCGGAAGGTGATCCT 4074
QY      167 LeuThrIleGlyValAlaAsnGluLeuThrHisGlyThrIleProProGlyLysIleThr 186
DB      4075 ATCACCGTGTGTGTAACACGAACTGAACCTGCGACATATCCCGCGGGAATGTGAT 4134
QY      187 ThrGlyAsnAlaThrGlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAla 206
DB      4135 ACC---GACGAAACCGCAAGAAAGACGCTTACTTCATCATTTCTTAACTATGCC 4191
QY      207 GlyLeuAlaArgSerIleTrrPleuTyrSerValProGlnGlnHisIleGlnAspIleThr 226
DB      4192 GGATCATCGACGCGTATGCTTACACCAACCGCCGAAACCTGGGGTGAACGATATCACC 4251
QY      227 ValValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGluValIleGluVal 244
DB      4252 GTGTGTGCGATGTGCGCAAGACTGTAACACCGCTGTGATCACTGGCAGCG---GTG 4308
QY      245 AlaAsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleVal 264
DB      4309 GCCAAT-----GGTGTGTGTCAGCGTTGAACCTGCGTATGCGATGCAACAGGTGTT 4359
QY      265 AlaAlaAlaSerGlyAlaGlnGlyThrValThrIleProSerValLysLeuTrrGlnPro 284
DB      4360 GCAACTGCAAGACGACTACCGGACTTTCGAACTGTGAATGCGCACCTCTGCAACCG 4419
QY      285 GlyAlaAlaIleTrrLeuTyrGlnLeuGlnValAlaAsnIleValGlySerSerGlyAspValVal 304
DB      4420 GGTAAGACTTATCTCTATGAACCTGTGCTC-----ACAGCAAAAGCCAGACGAGTGT 4473
QY      305 AspThrTyrAsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeu 324
DB      4474 GATATCTACCGCTTCGCGTCGCGCATCGGTCACTGCGAGTGAAGGCGCAACAGTTCTG 4533
QY      325 IleAsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGlyAspTrrAlaValArg 344
DB      4534 ATTACCAACCAACCGTTTACTTACTGTGGCTTGTGTGTGCTCAAGATGCGGACTTACGT 4593
QY      345 GlyLysGlyHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTrrIleGly 364
DB      4594 GCGAAGAGATTCGATTAACGTCGTATGTGTCACACCAACGATTAATGCACTGATGGG 4653
QY      365 AlaAsnSerPheArgThrSerHisTyrProTyrAlaGluGluValMetAspPheAlaAsp 384
DB      4654 GCCAACTCTACCGTACCTCCCATTAACCTTAACGTCGAAGAGTCTGCACTGGAGAT 4713
QY      385 ArgAsnGlyIleValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu--- 403
DB      4714 GAACATGGCATCTGTGTATGATGAACCTGCTGTGCGCTTAACTCTTTAGGC 4773
QY      404 MetGlyValSerGluSerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAsp 422
DB      4774 ATTGGTTTCGAAGCGCGCAACGCAAGCAAGAACTGTACAGCGAAGCGGTCAACCGG 4833
QY      423 LysThrGlnGluAlaHisIleGlyGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHis 442
DB      4834 GAAACTCAGCAAGCGCACTTACAGCGGATTAAGAGTGAATGCGCGTGAACAAACAC 4893

```

```

QY      443 AlaSerValValMetTrrSerIleAlaAsnGluProAlaSerHisGluAspGlyValArg 462
DB      4894 CCAAGCTGTGTGATGTGAGATTTGCCAAGAACCGGATCCCGTCT-----GCACGG 4947
QY      463 GluTyrPhe-----GluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgPro 480
DB      4948 GAATATTTCCGCAATTGCGCCACTGCGCGAAGCAACGGGTAACTCGAACCGCGTCCG 5007
QY      481 IleThrPheAlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPhe 500
DB      5008 ATCACCTGCGTCAACATGATATGTTCTGCGACCCCTACACCATACATACAGGATCTCTT 5067
QY      501 AspValSerCysIleAsnArgTyrPheGlyTrrPyrSerGlnThrGlyAspLeuGlu 520
DB      5068 GATGCGCTGTGCTGGAACCGTTATTAACGATGATGTCGCAAGCGCGATTTGGAAACG 5127
QY      521 AlaGluAlaAlaLeuGluLysGluLeuHisGlyTrrGlnGluLysPheHisArgProIle 540
DB      5128 GCAGAGAAAGTATCTGGAAAGAACTTCTGCGTGGCAGGAGAACTGCATCAGCCGATT 5187
QY      541 ValMetThrGluTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuPro 560
DB      5188 ATCATCACCAGATACCGCGGTGATACGTTACCGCGGTGCACTGAATGACACCGACATG 5247
QY      561 TrrSerGluGluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAspArgIle 580
DB      5248 TGAAGTGAAGAGTATCAGTGTGATGCTGATGATGTATCATCCGCTTTGATGTCGCTC 5307
QY      581 GluSerMetAlaGlyGluHisValIleTrrAsnPheAlaAspPheGlnThrAsnLeuGlyIle 600
DB      5308 ACGCGCGTGTGCGGAAACAGGTATGAATTTCCCGGATTTTGGCACTCCGAAAGGCTTA 5367
QY      601 IleArgValAspGlyAsnLysGlyValPheThrArgAspArgLysProLysAlaAla 620
DB      5368 TTGCGCGTTGGCGGTAAACAAGAAAGATCTTCACTCGCGAACCGAAACCGAATCGCG 5427
QY      621 AlaHisSerLeuArgAlaArgTrrPheSerIleAsp 632
DB      5428 GCTTTTCTGCTCAAAAACGCTGGAATGCGCATGAAC 5463

```

```

RESULT 12
US-09-434-039A-1
; Sequence 1, Application US/09434039A
; Patent No. 6531649
; GENERAL INFORMATION:
; APPLICANT: MANNERLOEF, Marie
; APPLICANT: TENNING, Paul Peter
; APPLICANT: STERN, Per
; TITLE OF INVENTION: Transgenic Plants
; FILE REFERENCE: 09/434,039
; CURRENT APPLICATION NUMBER: US/09/434,039A
; PRIOR FILING DATE: 1999-11-04
; PRIOR APPLICATION NUMBER: 60/112,003
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 09/182,117
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 8012
; TYPE: DNA
; ORGANISM: Sugar beet
US-09-434-039A-1

```

## Alignment Scores:

```

Pred. No.: 2,05e-171 Length: 8012
Score: 1692.00 Matches: 337
Percent Similarity: 68.35% Conservative: 95
Best Local Similarity: 53.32% Mismatches: 180
Query Match: 50.45% Indels: 20
DB: 4 Gaps: 11

```

US-10-757-093-4 (1-634) x US-09-434-039A-1 (1-8012)





```

Db      5407 GCTTTCTGCTGCAAAAACGCTGACTGGCATGAAC 5442
RESULT 14
US-09-434-039A-5
/ Sequence 5, Application US/09434039A
/ Patent No. 6531649
/ GENERAL INFORMATION:
/ APPLICANT: MANNERLOEF, Marie
/ APPLICANT: TENNING, Paul Peter
/ APPLICANT: STEEN, Per
/ TITLE OF INVENTION: Transgenic Plants
/ FILE REFERENCE: 09/434, 039
/ CURRENT APPLICATION NUMBER: US/09/434, 039A
/ CURRENT FILING DATE: 1999-11-04
/ PRIOR APPLICATION NUMBER: 60/112, 003
/ PRIOR FILING DATE: 1997-10-31
/ PRIOR APPLICATION NUMBER: 09/182, 117
/ PRIOR FILING DATE: 1998-10-29
/ NUMBER OF SEQ ID NOS: 36
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 5
/ LENGTH: 8418
/ TYPE: DNA
/ ORGANISM: Sugar beet
US-09-434-039A-5

Alignment Scores:
Pred. No.:      2,21e-171      Length:      8418
Score:          1692.00      Matches:      337
Percent Similarity: 68.35%      Conservative: 95
Best Local Similarity: 53.32%      Mismatches: 180
Query Match:    50.45%      Indels:      20
DB:             4           Gaps:          11

US-10-757-093-4 (1-634) x US-09-434-039A-5 (1-8418)
QY      10 LeuSerLeuAlaIaIaProSerLeuGlyThrProAlaAlaArgHisPheProArgAsnGlu 29
Db      3580 CTATCTCTTCGCAAGACCCCTTCTCTATATAGAGAGTTCATTTCATTTGGAGAGACACG 3639
QY      30 MetThrGlnHisGlnGlnProLeuLeuLeuValArgProGlnArgThrSerSerArgGlu 49
Db      3640 CTGACAGAGCTGACTAGCAGATCTCCATGTCCTGCTCTGTAAACCCCAACCCGGTGA 3699
QY      50 LeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu-----AlaSerGlyLeuAsn 66
Db      3700 ATCAAAAACCTCGACGGCGCTGTGGCATTCAGTCTGGATCGCAAAACTGTGGAATTGAT 3759
QY      67 AspThrAlaGlnProTrpThrAlaProLeuProLysGlyLeuGluCysProValProAla 86
Db      3760 CAG-----CGTTGGTGGAAAGCGCGTTACAGAAAGCCGGCAATTGCTGTGCCAGGC 3813
QY      87 SerTrpAsnAspIlePheIleSerArgGluIleHisAspHisValGlyTrpValTyrTyr 106
Db      3814 AGTTTAAAGATACAGTTCGCCGATGCGAGATTCGTAATTATCGGGCAACGTCGTGAT 3873
QY      107 GlnArgGluValIleValProLysGlyTrpSerGlnGluArgTyrLeuValArgAlaGlu 126
Db      3874 CAGCGGGAAGTCTTTATCCGAAAGGTTGGCAGCGCAGCGGTATCGTGGCGTTTCGAT 3933
QY      127 SerAlaThrHisIleGlyArgIleTyrValAsnAspArgLeuValAlaGluHisValGly 146
Db      3934 GCGGTACATCATTAAGCAAAAGTGGGATTAATTAACAGAAAGTGAATGAGCATCAAGGC 3993
QY      147 GlyTrpThrProPheGluAlaAspValThrGluLeuValAlaProGlyGluLysPheArg 166
Db      3994 GGGTAAACCGCATTTGAAGCCGATGTCACGCCGATGTTATTCGCCGGAAAGATGTAAGT 4053
QY      167 LeuThrIleGlyValAsnAsnGlyLeuThrHisGluThrIleProProGlyLysIleThr 186
Db      4054 ATCACCGTTTGTGTGAACAAAGAACTGAATCGGCAGACTATCCCGCGGGAATGTGATTT 4113

```

```

QY      187 ThrGlyAsnAlaThrGlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAla 206
Db      4114 ACC---GACGAAAGCGCAAGAAAGACGCTTACTTCATCATATTCTTTAACTATGCC 4170
QY      207 GlyLeuAlaArgSerIleTrpLeuTyrSerValProGlnGlnHisIleGlnAspIleThr 226
Db      4171 GGAATCATTCGACGCTAATGCTCTACACCAACCCGAAACCTGGGGGACGATATCACC 4230
QY      227 ValValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGluValGluVal 244
Db      4231 GGTGTACCGATGTCGCGCAAGCTGTACCAACCGCTCTGTATCGGCAGGTC---GTG 4287
QY      245 AlaAsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleVal 264
Db      4288 GCCAAT-----GCTATGTCACCGCTTGAACCTGCGATCCGATCAACAGTGTGTT 4338
QY      265 AlaLysAlaSerGlyValaGlnGlyThrValThrIleProSerValLysLeuTrpGlnPro 284
Db      4339 GCAACTGCAACAGCGACTAGCGGACTTTCGAAGTGTGAATCCGACCTCTGGCAACCG 4398
QY      285 GlyAlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValVal 304
Db      4399 GGTGAAGTTATCTCTATGAACCTGTGCTC-----ACAGCCAAAAGCCAGACAGTGT 4452
QY      305 AspThrTyrAsnLeuAlaThrGlyValArgThrValLysValaIaGlySerGlnPheLeu 324
Db      4453 GATATCTACCCGCTTTCGCGTCGCGCATCCGTCAGTGGCAAGGCGCAACAGTTCCTG 4512
QY      325 IleAsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGlyAspThrAlaValArg 344
Db      4513 ATTAACCAACAAACGTTTACTTACTTACGCTTGGCTGTGCTCATGAAGTCGCGACTTAAGT 4572
QY      345 GlyLysGlyHisAspProAlaTyrMetValHisAspPheIleuMetLysTrpIleGly 364
Db      4573 GGCAAAGGATTCATTAACGCTGATGTGACAGACCAACGCAATTAATGACTGATTTGGG 4632
QY      365 AlaAsnSerPheArgThrSerHisTyrProTyrAlaGlyGluValMetAspPheAlaAsp 384
Db      4633 GCCAACTCTCAACGTCATCTTCGATTAACCTTACGCTGAAGAGATTCCTCGACTGGCAGAT 4692
QY      385 ArgAsnGlyTyrLeuValIleAspGlyThrProAlaValGlyLeuAsnIleAlaLeu--- 403
Db      4693 GAACATGATGATCGTGTGATGATGAACGCTGCTGCGCTTTAACTCTCTTTAAGC 4752
QY      404 MetGlyValSerLysSerGlyAlaPro---GlnThrPheThrProAspAlaIleAspAsp 422
Db      4753 ATTGGTTTGGAAACGGGCAACAGCCGAAAGAACTGTACAGCAAGAGAGTCAACGGG 4812
QY      423 LysThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHis 442
Db      4813 GAACCTCAGACAGCGCATTTACAGGCGATTAAAGACTGATACCGCTGCACAAACAC 4872
QY      443 AlaSerValValMetTrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArg 462
Db      4873 CCAAGCGGTGATGTGAGATATTGCCAAGCAACCGGATACCGTCT-----GCACGG 4926
QY      463 GlyTrpPhe-----GluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgPro 480
Db      4927 GAATATTTTGGCATTTTCGCCACTGGCGGCAAGCAACCGTAACTTCACCGCAGCGCTCCG 4986
QY      481 IleThrPheAlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPhe 500
Db      4987 ATCACTGCGTCATTAATGTTCTGCGAGCTCAACCGATACATCAGCGATCTCTTT 5046
QY      501 AspValSerCysIleAsnArgTyrPheGlyTyrTrpSerGlnThrGlyAspLeuGluGlu 520
Db      5047 GATGTCTGTCGCTGAACCGTTATTAACGATGTATGTCCAAAGCGCGGATTTGAAACG 5106
QY      521 AlaGluAlaAlaLeuGluLysGlyLeuHisGlyTyrGlnGluLysPheHisAspProIle 540
Db      5107 GCAGAGAAAGTACTGGAAAGAACTTGGCGCTGCAGAGAAAGTGCATGACGGCATTT 5166
QY      541 ValMetThrGluTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuPro 560

```

```

Db      5167 ATCATCACGAAATACGCGTATAGCTTACCGCGGTGCATCTCAATGTACACCGACATG 5226
Qy      561  TTPSerTInGluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAspArgTle 580
Db      5227 TGGAGTAAAGATACGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 5286
Qy      581  GluSerMetAlaGlyGlnHisValTTPAsnPheAlaAspPheGlnThrAsnLeuGlyTle 600
Db      5287 AGCGCCGCTGCTGCTGTAACAGTATGTAATTTCCGCCATTTTGGACCTCGCAAGGCATA 5346
Qy      601  IleArgValAspGlyAsnLysLysGlyValPheThrArgAspArgLysProLysAlaIle 620
Db      5347 TTGGCGCTTGGCGGTACCAAGAAAGGATCTTCACGCGCAACCGAAAGTGGCGG 5406
Qy      621  AlaHisSerLeuArgAlaArgTTPThrSerIleAsp 632
Db      5407 GCTTTTGTCTGCAAAAACGCTGGACTGGCATGTAC 5442

```

```

RESULT 15
US-09-182-117-4
/ Sequence 4, Application US/09182117
/ Patent No. 6204436
/ GENERAL INFORMATION:
/ APPLICANT:
/ TITLE OF INVENTION: Transgenic Plant
/ NUMBER OF SEQUENCES: 27
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: IBM PC compatible
/ SOFTWARE: Patent Release #1.0, Version #1.25 (BPO)
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/182,117
/ FILING DATE:
/ CLASSIFICATION:
/ INFORMATION FOR SEQ ID NO: 4:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 8798 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ HYPOTHEICAL: NO
/ ANTI-SENSE: NO
/ US-09-182-117-4

```

```

Alignment Scores:
Pred. No.: 2,388-171 Length: 8798
Score: 1692.00 Matches: 337
Percent Similarity: 68.35% Conservative: 95
Best Local Similarity: 53.32% Mismatches: 180
Query Match: 50.45% Indels: 20
DB: 3 Gaps: 11

```

US-10-757-093-4 (1-634) x US-09-182-117-4 (1-8798)

```

Qy      10  LeuSerLeuAlaAlaProSerLeuGlyThrProAlaAlaArgHisPheProArgGlnGlu 29
Db      3690 CTATCTCTGCAAGACCTTCTCTATATAGAAGATTCAATTTCATTGGAGAGACAG 3749
Qy      30  MetThrGlnHisGlnGlnProLeuLysLysValArgProGlnArgThrSerSerArgGlu 49
Db      3750 CTGACAAAGTGACTTAGAGAGATCTTCATGTCCTGTAAGAAACCCCAACCCGCTGA 3809
Qy      50  LeuValAsnLeuAspGlyLeuTTPlysPheAlaLeu-----AlaSerGlyLeuAsn 66
Db      3810 ATCAAAAACCTGACGCGCTGAGGATTCAGCTGGATCGGAAACCTGTGAATGTAGT 3869
Qy      67  AspThrAlaGlnProTTPThrAlaProLeuProLysGlyLeuGlnCysProValProAla 86
Db      3870 CAG-----CGTTGGTGGAAAGCGGCTTACAAAGAAAGCGGCAATTGCTGTGCGAGCG 3923

```

```

Qy      87  SerTyrAsnAspIlePheIleSerArgGluIleHisAspHisValGlyThrValTyrTyr 106
Db      3924 AGTTTAACGATAGTATGTCGCCGATGCAAGATATTCGTAATTAAGCGGCAACGCTGTGAT 3983
Qy      107  GlnArgGluValIleValProLysGlyTTPSerGlnLysArgTyrLeuValArgAlaGlu 126
Db      3984 CAGCGCAAGTCTTTATACCGAAAGGTTGGGACGCGCAAGATGATGCTGCTGCTTCAT 4043
Qy      127  SerAlaThrHisHisGlyArgLysTyrValAsnAspArgLeuValAlaGlnHisValGly 146
Db      4044 GCGGTACATCATTAATCGCAAGAGTGGTCAATTAATACAGGAAGTGTAGAGATCAGGGC 4103
Qy      147  GlyTyrThrProPheGlnAlaAspValThrGluLeuValAlaProGlyGlnLysPheArg 166
Db      4104 GGTATACCCCATTTGAACCCATGTCACGCGGTATTTATGGCGGAAAGTGTACCT 4163
Qy      167  LeuThrIleGlyValAsnAsnGlnLeuThrHisGlnThrIleProProGlyLysIleThr 186
Db      4164 ATCACCGTTTGTGTGAACAACGAACGAACTGAGACTATATCCCGCGGAAATGTGATT 4223
Qy      187  ThrGlyAsnAlaThrGlyLysArgLysGlnThrTyrGlnHisAspPheTyrAsnTyrAla 206
Db      4224 ACC--GACGAAACCGCAAGAAAGAGCTCTTACCTTCATGATTTCTTTAATCTATGCC 4280
Qy      207  GlyLeuAlaArgSerIleTTPLeuTyrSerValProGlnGlnHisIleGlnAspIleThr 226
Db      4281 GGATCCATCGCAGGTATAGCTTACACCGCCGAACACCTGGGTGACATATCAC 4340
Qy      227  ValValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGluValGluVal 244
Db      4341 GTGTGACCGCATGTGCGCGCAAGCTGTAAACCGGCTGTGTGATGCGGACGTG--GTG 4397
Qy      245  AlaAsnGlnThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleVal 264
Db      4398 GCCAAT-----GGTATGTACCGTTGACCTGCGTATGCGGATCAACAGGTGTT 4448
Qy      265  AlaLysAlaSerGlyAlaGlnGlyThrValThrIleProSerValLysLeuTTPGlnPro 284
Db      4449 GCACTGACCAAGGACGTAAGCGGACTTGAAGTGGTAATCCGACCTCGGCAACGG 4508
Qy      285  GlyAlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValVal 304
Db      4509 GGTGAAGTTATCTTATGAACTGTGCGTC-----ACAGCCMAAGCCAGACAGTGT 4562
Qy      305  AspThrTyrAsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeu 324
Db      4563 GATATCTACCGCTTCGCGGTGCGGACATCCGCTACAGTGAAGGCAAGACATGCTCTG 4622
Qy      325  IleAsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGlnAspThrAlaValArg 344
Db      4623 ATTAACCAACAAACCGTTCTACTTACTGCTTGTGCTCATGAAGATGCGGACTTACGT 4682
Qy      345  GlyLysGlnHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTTPIleGly 364
Db      4683 GGCAAAGATTCGAATACGCTGATGTGACAGCAACGCAATTAATGACTGTGATTTGGG 4742
Qy      365  AlaAsnSerPheArgTyrSerHisTyrProTyrAlaGlnGluValMetAspPheAlaAsp 384
Db      4743 GCGAATCTCTACCGTACCTCGCATTAACCTTACGCTGAAGATGCTGATGCGGACAT 4802
Qy      385  ArgAsnGlyIleValValIleAspGlnThrProAlaValGlyLeuAsnIleAlaLeu--- 403
Db      4803 GAACATGCGATCGTGTGATGATGAACCTGCTGCGCTTAACTCTCTTTAGGC 4862
Qy      404  MetGlyValSerGlnSerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAsp 422
Db      4863 ATTGTTTCGAAGCGGCGCAACAGCCGMAAGACTGTACACGGAAGACGCTCAACGGG 4922
Qy      423  LysThrGlnGlnAlaHisLysGlnAlaIleArgGlnLeuIleAlaArgAspLysAsnHis 442
Db      4923 GAACCTGCAAGAGCGCACTTACAGCGCATTAAGAAGCTGATAGCGCGGACAAACACAC 4982
Qy      443  AlaSerValValMetTTPSerIleAlaAsnGlnProAlaSerHisGlnAspGlyAlaArg 462

```

```
Db 4983 CCAAGCGTGGTATGAGATATTGCCAAGAACCGGATACCGCTCT-----GCACGG 5036
Qy 463 GluTyrPhe-----GluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgPro 480
Db 5037 GAATATTTCGGCATTTCCGCACTGGCGAAGCAACCGGTAACTCGACCCGACCGCGTCCG 5096
Qy 481 IleThrPheIleAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPhe 500
Db 5097 ATACCTGCGCTCAATGTAATGTTCTGCGACGCTCACACCGATACCAATCAGCAATCTCTTT 5156
Qy 501 AspValSerCysIleAsnArgTyrPheGlyTyrTyrSerGlnThrArgIleAspLeuGlu 520
Db 5157 GATGCTGTGCTGCTGAACCGTTATTCAGATGGATGTCCAAAGCGCGATTTGGAAACG 5216
Qy 521 AlaGluAlaAlaLeuGluLysGluLeuHisGlyTyrGlnGluLysPheHisArgProIle 540
Db 5217 GCAGACAAAGGTACTGGAAAAAAGAACTTCTGCGCGAGAGAAACTGCATCAGCCGATT 5276
Qy 541 ValMetThrGluTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuPro 560
Db 5277 ATCATCCGCAATACCGCGCTGCAATACGTTACCGGCGCTGCATCATGTAACCGACATG 5336
Qy 561 TrpSerGluGluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAspArgIle 580
Db 5337 TGGAGTGAAGAGTATCAGTGTGATGCTGATGATATGATCACCGCGCTTTGATCGGCTC 5396
Qy 581 GluSerMetAlaGlyGluHisValTyrAsnPheAlaAspPheGlnThrAsnLeuGlyIle 600
Db 5397 AGCGCGCTGCTCGGTAAACAGTATGGAATTTGCGCCGATTTGCGACTCGCAAGGCATA 5456
Qy 601 IleArgValAspGlyAsnLysLysGlyValPheThrArgAspArgLysProLysAlaAla 620
Db 5457 TTCGGCGTTGGCGTAACAAGAAAGGATCTTCACTGCGAACCGCAACCGAAGTCGGCG 5516
Qy 621 AlaHisSerLeuArgAlaArgTyrThrSerIleAsp 632
Db 5517 GCTTTTCTGCTGCATAAAGCGTGCATGATGAC 5552
```

Search completed: March 22, 2005, 20:44:17  
Job time : 407 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM protein - nucleic search, using frame\_p2n model

Run on: March 22, 2005, 14:08:15 ; Search time 781 Seconds  
(without alignments)  
4805.526 Million cell updates/sec

Title: US-10-757-093-4  
Perfect score: 3354  
Sequence: 1 MKFLTGLSLTSLAPSLGTP.....RKPKAAASHLRARWTSIDKN 634

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 segs, 2959870667 residues  
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seg length: 0  
Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

-MODE=frame\_p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool\_p/US10757093/runat\_18032005\_164455\_27758/app\_query.fasta\_1.775  
-DB=N.Geneseq\_16Dec04 -QPM=faeta -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bloms62 -TRANS=human40.cdi  
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTMT=ptc -NCRM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10757093.@CGN\_1.1.708.@runat\_18032005\_164455\_27758 -NCPD=6 -ICPD=3  
-NO\_MMAP -LARGESUBERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

1: N.Geneseq\_16Dec04:.\*  
2: geneseqn1980s:.\*  
3: geneseqn1990s:.\*  
4: geneseqn2000s:.\*  
5: geneseqn2001s:.\*  
6: geneseqn2002as:.\*  
7: geneseqn2002bs:.\*  
8: geneseqn2003as:.\*  
9: geneseqn2003bs:.\*  
10: geneseqn2003cs:.\*  
11: geneseqn2003ds:.\*  
12: geneseqn2004as:.\*  
13: geneseqn2004ds:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1756	52.4	1812	3	AAA07939
2	1717.5	51.2	3451	8	ABT16609
3	1717.5	51.2	3451	10	ACC44711
4	1717.5	51.2	14627	8	ABT16610
5	1717.5	51.2	14627	10	ACC44712

6	1717	51.2	4299	3	AAA27980	AAa27980 Vector MS
7	1714.5	51.1	3169	2	AA138397	Aa138397 pTc 99 p
8	1714.5	51.1	8654	4	AA180288	Aa180288 Nucleotid
9	1714.5	51.1	8654	4	AA180285	Aa180285 Nucleotid
10	1714.5	51.1	9390	4	AA180296	Aa180296 Nucleotid
11	1714.5	51.1	9390	4	AA180292	Aa180292 Nucleotid
12	1714.5	51.1	9390	4	AA180294	Aa180294 Nucleotid
13	1714	51.1	5919	4	AA180980	Aa180980 pSP-GUS
14	1714	51.1	7510	12	AD101287	Ad101287 Plasmid p
15	1714	51.1	8327	12	AD101291	Ad101291 Plasmid p
16	1714	51.1	10011	12	AD101297	Ad101297 Plasmid p
17	1714	51.1	12225	6	AD101298	Ad101298 Plasmid p
18	1714	51.1	12225	6	AD101299	Ad101299 Plasmid p
19	1714	51.1	32798	6	AB16684	Ab16684 Replicati
20	1714	51.1	32798	6	AB16684	Ab16684 Replicati
21	1713.5	51.1	6795	3	AA166821	AA166821 Reporter
22	1713.5	51.1	6795	3	AA166822	AA166822 Reporter
23	1713.5	51.1	6795	3	AA166817	AA166817 Reporter
24	1712.5	51.1	3035	2	AA137748	Aa137748 PAT1 Gene
25	1711	51.0	1809	6	AB167080	Ab167080 uida DNA
26	1711	51.0	1812	6	AB161371	Ab161371 GUS DNA #
27	1711	51.0	1812	6	AB161371	Ab161371 GUS DNA #
28	1711	51.0	1812	8	AB161371	Ab161371 GUS DNA #
29	1711	51.0	1812	10	ACC44709	Acc44709 E. coli b
30	1711	51.0	1812	13	AD166099	Ad166099 Bacterial
31	1709	51.0	3824	2	AA137749	Aa137749 PAT1 gene
32	1709	51.0	1812	4	AA122503	Aa122503 PCR gene
33	1709	51.0	1812	6	AA129543	Aa129543 uida DNA
34	1709	51.0	4652	10	ADD27985	Add27985 Beta-gluc
35	1709	51.0	8617	3	AA166819	AA166819 Reporter
36	1709	51.0	8617	3	AA166818	AA166818 Reporter
37	1706	50.9	5390	3	AA166820	AA166820 Reporter
38	1706	50.9	5418	10	ADD27987	Add27987 Oleosin/b
39	1703	50.8	11978	6	AB153598	Ab153598 Neureoth
40	1703	50.8	12438	6	AB153599	Ab153599 Plasmid p
41	1700	50.7	1814	10	AD101286	Ad101286 Enterohae
42	1697	50.6	1812	6	AB161358	Ab161358 GUS DNA #
43	1692	50.4	8012	2	AA157305	Aa157305 Sugar bee
44	1692	50.4	8012	2	AA157309	Aa157309 Sugar bee
45	1692	50.4	8798	2	AA157308	Aa157308 Sugar bee

## ALIGNMENTS

RESULT 1	AAA07939	standard; DNA; 1812 BP.
ID	AAA07939	
XX	AAA07939	
AC	AAA07939	
DT	26-JAN-2001	(first entry)
XX		
DE	Salmonella beta-glucuronidase gene.	
XX		
KW	Microbial; beta-glucuronidase; GUS; Enterobacter; Salmonella;	
KW	Pseudomonas; Staphylococcus; Thermotoga; transgenic plant; bioindicator;	
KW	transgenic insect; marker; glucuronide detoxification; ds.	
XX		
OS	Salmonella sp.	
XX		
PN	WO20005533-A1.	
XX		
PD	21-SEP-2000.	
XX		
PF	16-MAR-2000; 2000WO-US007107.	
XX		
PR	17-MAR-1999; 99US-00270957.	
XX		
PA	(CAMP-) CAMBIA BIOSYSTEMS LLC.	
XX		
PI	Jefferson RA, Mayer JE;	
XX		
DR	WPI; 2000-647075/62.	

DR P-PSDB; AAB28409.

XX Novel microbial beta-glucuronidase genes and gene products used as  
PT reporter/effector molecule, as diagnostic tool, in positive selection, to  
PT target molecules to specific cells and to detect and track linked genes.

XX Example 3; Fig 16; 116pp; English.

CC The present sequence encodes a microbial beta-glucuronidase (GUS)  
CC protein. GUS genes were obtained from six different genera:  
CC Enterobacter/Salmonella, Pseudomonas, Salmonella, Staphylococcus and  
CC Thermotoga. Microbial GUS can be used as a reporter/effector molecule for  
CC transgenic constructions and in in vitro diagnostic applications. It may  
CC also be used to generate sentinel plants that serve as bioindicators of  
CC environmental status. It may be used to facilitate the development of a  
CC tracking insect populations or to facilitate the development of insect  
CC bioassay for compounds that affect molecules critical for insect  
CC development (e.g. juvenile hormone). Secreted GUS may also serve as a  
CC marker for beneficial fungi destined for release into the environment. In  
CC animal systems, secreted GUS may be used to achieve extracellular  
CC detoxification of glucuronides (e.g. toxin glucuronide) and to examine  
CC conjugation patterns of glucuronides. Microbial GUS may also be used in  
CC traditional medical diagnostic assays, for drug testing, pharmacokinetic  
CC studies, bioavailability studies, diagnosis of diseases and syndromes,  
CC following progression of disease or its response to therapy. Microbial  
CC GUS has increased thermal stability, high turnover number and enzymatic  
CC activity. It is highly specific for the substrate and water soluble, and  
CC the substrates are stable

XX Sequence 1812 BP; 437 A; 465 C; 531 G; 379 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1,21e-142	Length:	1812
Score:	1756.00	Matches:	329
Percent Similarity:	71.6%	Conservative:	99
Best Local Similarity:	55.11%	Mismatches:	159
Query Match:	52.36%	Indels:	10
		Gaps:	6

US-10-757-093-4 (1-634) x AAA07939 (1-1812)

```

QY 40 ValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpLysPhe 59
DB 4 TTTACGTTCTGTCGAAACCGCGAGCGCGAATCAAAAACCTGACCGCTGTGGTGGTTT 63
QY 60 AlaLeuAspGlyLeuAsnAspThrAlaGlnPro---TrpThrAlaProLeuProLys 78
DB 64 TGTATGATAGCGAAGATCGCGCAACGCGACGATGTCGCTCAACCGTTAACCCCAA 123
QY 79 GlyLeuGlycysProValProAlaSerThrAsnAspIlePheIleSerArgGluIleHis 98
DB 124 AGCGCGCCTATCCGCGTTCCGGAGAGCTATACATGATGTTGCCGTCGCGAGATCCG 163
QY 99 AspHisValGlyTrpValAlaGlyValGlyValAlaValProLysGlyTrpSerGln 118
DB 184 AATTATGTGGCAACGCTGCTGATACCGTGAGATAGCATCCGGAAGCTGGAGATCC 243
QY 119 GluArgTrpLeuValAlaGlyAlaGlySerAlaThrHisIleGlyArgGlyIleValAsnAsn 138
DB 244 CAGCGCATAGTGTGCGCTTGTGATGCGTACACTACATGAGAAAGTTTGGTCAATGAC 303
QY 139 ArgLeuValAlaGlyValGlyValGlyValGlyValGlyValGlyValGlyValGlyVal 158
DB 304 CAATTTTATATGAGAACATCAGGGCGGTACACGCGTTTGAGGGGATATACGACACTT 363
QY 159 ValAlaProGlyGlyValGlyValGlyValGlyValGlyValGlyValGlyValGlyVal 178
DB 364 ATCTCCCGCGGAGATCCGTCGTATACCGTATGCGTATGATGATGATGATGATGATGATG 423
QY 179 ThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysArgGlyIleThrThr 198
DB 424 ACGATCCCGCGCGGCTGTTGTGACCAAGGC---GTAAACGGTAAAGACAGCAAGCCGTAT 480

```

```

QY 199 GlnHisAspPheThrAsnTrpAlaGlyLeuAlaArgSerIleTrpLeuTrpLysValPro 218
DB 481 TTCATGATTTTCTTACTACCGCGTATTCATTCGACGATGATGATGATGATGATGATGATG 540
QY 219 GlnGlnHisIleGlnAspIleThrValValThrAspValAspGlyAspAsnGlyLeuIle 238
DB 541 AAAACCTTTGTGAGATATTACCTGTCGACGCGAGGTGTCAGCAT-----CTGGCT 594
QY 239 AsnTrpGlyValGlyValAlaAsnGlnThrThrGlnIleGlnIleSerValIleAsp 258
DB 595 CAGGCTACCGTCCCTGCGAGTACGCGGAGTAAGCGAATGCGAATGCTGTAGCTTACGATG 654
QY 259 GluAspGlyAlaIleValAlaLysAlaSerGlyAlaGlnGlyValThrValThrIleProSer 278
DB 655 GCGAGCAACACACTTGTGCTTCGCGGCAAGGGAAGAAAGTGAACCTGCTGCTGAGAGG 714
QY 279 ValLysLeuTrpGlnProGlyAlaAlaTrpLeuTrpGlnLeuGlnValAsnIleValGly 298
DB 715 CCGCGCTGTGCGAGCTGCGAGGCTATCTTTATGAACTCGGGTC-----ATCGCG 768
QY 299 SerSerGlyAspValValAspThrTrpAsnLeuAlaThrGlyValArgThrValLysVal 318
DB 769 CAGCATCAGACGACGAGATGAATATCCGCTGCGCGTGGTATTCGCTCGGTAGAGTA 828
QY 319 AlaGlySerGlnPheLeuIleAsnGlyLysProPheTrpPheThrGlyPheGlyLysHis 338
DB 829 AAAGGGACACATGCTCTGATCAACCATTAAGCTTTCTATTATTCACCGGGTTCGACGTAT 888
QY 339 GluAspThrAlaValArgGlyLysGlyHisAspProAlaTrpMetValHisAspPheGln 358
DB 889 GAAGATCCGACATCGCGCGTAAAGGTTTGTATTAACGTGATGATGATGATGATGATGATG 948
QY 359 LeuMetLysTrpIleGlyValAsnSerPheArgThrSerHisIleTrpTrpAlaGlnGlu 378
DB 949 CTATATGACTGTGATCGTGCAGACTTCTTACCGTACCTGACATTCACCTTATTCGCAAGAG 1008
QY 379 ValMetAspPheAlaAspArgAsnGlyIleValIleAspGlyLysProAlaValGly 398
DB 1009 ATGCTGACATGGGCGGACCAACATGATGATGATGATGATGATGATGATGATGATGATG 1068
QY 399 LeuAsnIleAlaLeuMetGlyValSer-----GluSerGlyValArgGlnThrPhe 415
DB 1069 TTCACATCTGCTTTA---GGGATTCATGCTTGTATGTCGCGGAAACCCAAAGAGCTTAC 1125
QY 416 ThrProAspAlaIleAsnAspLysThrGlnGlnAlaHisIleGlyAlaIleArgGlnLeu 435
DB 1126 AGCGATGAGCGCTGTAACCATTAAGACGCGCGGACCTGACGACATTAAGAGAGCTG 1185
QY 436 IleAlaArgAspLysAsnHisAlaSerValAlaMetTrpSerIleAlaAsnGluProAla 455
DB 1186 ATTGCCCGGATTAAGAACCAACCAACGCTGATGATGATGATGATGATGATGATGATGATG 1245
QY 456 SerHisGluAspGlyAlaArgGlyLysPheGluProLeuThrAsnLeuThrArgGlnLeu 475
DB 1246 ACCGCGCCGAACGCGCGCGCAATCTTGCCTGCGCTGCGCGCAAGCAACGCGCAATCTC 1305
QY 476 AspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTrpGlnLeuAspArg 495
DB 1306 GATCTTACACGCTCGATACCTCGCGAACGATGATGATGATGATGATGATGATGATGATG 1365
QY 496 IleSerAspLeuPheAspValSerCysIleAsnArgTrpPheGlyTrpTrpSerGlnThr 515
DB 1366 ATTACCGATCTCTTATGTCGTTTTCGTAACCGCTACCGGCTGATGATGATGATGATG 1425
QY 516 GlyAspLeuGlnGluAlaGluAlaAlaLeuGluLysGlnLeuHisIleGlyTrpGlnGluLys 535
DB 1426 GCGGATCTGAGAGAGCTGAGAAAGTCTGAGAAAGAACTTCTGCGCGGAGGAGAA 1485
QY 536 PheHisArgProIleValMetThrGluTrpGlyAlaAspThrLeuAlaGlyLeuHisSer 555
DB 1486 CTCACCGCGCGGCTTATCATCACCGAATACGAGGCTGATGATGATGATGATGATGATG 1545
QY 556 IleLeuGlyLeuProTrpSerGlnGluGluGlnValGlnMetLeuAspMetTrpHisArg 575

```

Db 1546 ATGTACACGATATGTGACGGAAGATACCAAGCGCTGCTGATATGATCCATCCG 1605  
Qy 576 ValPheAspArgIleGluSerMetAlaGlyGluHisValTrpAsnPheAlaAspPheGln 595  
Db 1606 GTCTTTGATGCGCTGACGCGCGCTGCGGACAGGATTAAGAACTTCGCCGCTCGCC 1665  
Qy 596 ThrAsnLeuGlyIleIleArgValAspGlyAsnLysLysGlyValPheThrArgAspArg 615  
Db 1666 ACTTCGACGGGCTTATGCGCGCTTGGCGCAAAAAGCTATATTCCACCCCGACAGA 1725  
Qy 616 LysProLysAlaAlaAlaHisSerLeuArgAlaArgTrpHisSerIleAsp 632  
Db 1726 AAACCAAAATCGCGCGCTTCCTGCTGCAAAAACGCTGACCGCATGAC 1776  
RESULT 2  
ID ABT16609 standard; DNA; 3451 BP.  
XX AC ABT16609;  
XX DT 03-APR-2003 (first entry)  
XX DE Artificial plant chromosome related plasmid DNA SEQ ID No 20.  
XX KM Plant artificial chromosome; PAC; transgenic plant; vaccine;  
XX KM blood factor; herbicide; stress; agronomical; nutrient quality;  
XX KM bacterial artificial chromosome; BAC; yeast artificial chromosome; YAC;  
XX OS ds.  
XX Unidentified.  
XX PN MO200296923-A1.  
XX PD 05-DEC-2002.  
XX PF 30-MAY-2002; 2002MO-US017451.  
XX PR 30-MAY-2001; 2001US-0294687P.  
XX PR 04-JUN-2001; 2001US-0296329P.  
XX PA (CHRO-) CHROMOS MOLECULAR SYSTEMS INC.  
XX PA (AGRI-) AGRISOMA INC.  
XX PI Perez C, Fabijaneki SF, Perkins E;  
XX WPI; 2003-140436/13.  
XX PT Producing artificial chromosome by introducing a nucleic acid into plant  
XX PT cell, selecting artificial chromosome that has one or more repeat regions  
XX PT with equivalent amounts of euchromatic and heterochromatic nucleic acids.  
XX PS Disclosure; Page 248-249; 2699p; English.  
XX The invention relates to a novel method for producing plant artificial  
XX chromosomes. The invention also relates to methods for targeting  
XX insertion of heterologous DNA into plant artificial chromosomes, methods  
XX for delivery of plant chromosomes to selected cells and tissues. The  
XX isolated plant artificial chromosome (PAC) is useful for producing a  
XX transgenic plant, which involves introducing the PAC into a plant cell.  
XX The PAC comprises a heterologous nucleic acid encoding a gene product  
XX such as enzymes, antisense RNA, rRNA, rdna, structural proteins, marker  
XX proteins, ligands, receptors, ribozymes, therapeutic proteins, and  
XX biopharmaceutical proteins, vaccines, blood factors, antigens, hormones,  
XX cytokines, growth factors, antibodies, or a product that provides for  
XX resistance to diseases, insects, herbicides, or stress in a plant. The  
XX heterologous nucleic acid optionally encodes a product that provides an  
XX agronomically important trait in the plant, e.g. a product that alters  
XX nutrient use and/or improves the nutrient quality of the plant. The  
XX heterologous nucleic acid is contained within a bacterial artificial  
XX chromosome (BAC) or a yeast artificial chromosome (YAC). This  
XX polynucleotide sequence represents the DNA of a plasmid used in the  
XX method of the invention

XX SQ Sequence 3451 BP; 951 A; 707 C; 851 G; 942 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 6,15e-139 Length: 3451  
Score: 1717.50 Matches: 338  
Percent Similarity: 68.84% Conservative: 95  
Best Local Similarity: 53.74% Mismatches: 175  
Query Match: 51.21% Indels: 21  
DB: Gaps: 11  
US-10-757-093-4 (1-634) x ABT16609 (1-3451)  
Qy 13 AlaAlaProSerIleuGlyThrProAlaAlaArgHisPheProArgAsnGlu----- 29  
Db 1221 TCTGCACCGGATTCAGATC-----GAATTCCTCGCGCGCGGCAATTCAGTAC 1268  
Qy 30 ---MetThrGlnHisGluGlnProLeuIleLysValArgProGlnArgThrSerSerArg 48  
Db 1269 TGGATCCCGCGGATCGGATCGATTCCTTATG---TTACGTCCTGTAGAAACCCCAACCGGT 1325  
Qy 49 GluLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu---AlaSerGlyLeuAsnAsp 67  
Db 1326 GAAATCAAAAACCTCGACGCGCTGTGGCAATTCAGTCTGATGCGCAAAACCTGGAATT 1385  
Qy 68 ThrAlaGlnProTrpThrAlaProLeuProLysGlyLeuGluCysProValProAlaSer 87  
Db 1386 GAGCAGCGTTGGTGGGAAAGCGGCTTACAAAGAACCGGCAATTCGTCTGCAAGCAGT 1445  
Qy 88 TyrAsnAspIlePheIleSerArgGluIleHisAspHisValGlyTrpValYTrpYgin 107  
Db 1446 TTTAAAGATCAGTTCGCCGCGATGCAATTCGAAATTAATGTTGGCAACGCTGATACAG 1505  
Qy 108 ArgGluValIleValProLysGlyTrpSerGlnGluArgGlyLeuValArgAlaGluSer 127  
Db 1506 CGCGAAGCTCTTATACCGAAAGGTTGGGCAAGCGCATCTGCTGCTTCGATGCG 1565  
Qy 128 AlaThrHisHisGlyIleGlyValLeuAsnAspArgLeuValAlaGluHisValGly 147  
Db 1566 GTCACTCATTTACCGCAAGATGTTGGTCAATTAATCAGAAAGTGTGAGCATCAGGCGGC 1625  
Qy 148 TyrThrProPheGlnAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeu 167  
Db 1626 TATAGCCCATTTGAACCCGATGTCACGCGTATGTTATGCGGAAAGTGAACGTATC 1685  
Qy 168 ThrIleGlyValAsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThr 187  
Db 1686 ACAATTGTGTGAACAACGAACTGAACCTGACAGACTATCCCGCGGAAATGATGATACC 1745  
Qy 188 GlyAsnAlaThrGlyLysArgIleGlnThrTrpGlnHisAspPheTrpAsnTrpAlaGly 207  
Db 1746 ---GACGAAAACCGCAAGAAAAGACGCTTACTTCCATGATTTCTTAACCTAGCCGGG 1802  
Qy 208 LeuAlaArgSerIleTrpLeuTrpYserValProGlnGlnHisIleGlnAspIleThrVal 227  
Db 1803 ATCCATCGCAGCGTATATGCTTACACACCGCAACACCTGGGTGACATACACCGTG 1862  
Qy 228 ValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTrpGluValAla 245  
Db 1863 GTGACCAAGTCCGCGCAAGCTGTACACACGCTGTTGATCGACGAGG---GTGGCC 1919  
Qy 246 AsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAla 265  
Db 1920 AAT-----GATATGACCGCTTGAACCTCGTAATCGGATCAACAGGTGTGCA 1970  
Qy 266 LysAlaSerGlyValaGlnGlyThrValThrIleProSerValLysLeuTrpGlnProGly 285  
Db 1971 ACTGACAAAGCAGCAGCGGAGCTTTCGAAAGTGTGAATCCGCACTTCGCAACGGGGT 2030  
Qy 286 AlaAlaTrpLeuTrpGlnLeuGlnValAsnIleValGlySerSerGlyAspValValAsp 305  
Db 2031 GAAGTTATCTTATGAACTGATGTC-----ACAGCAAAAGCCAGACAGAGTGTAT 2084

```

QY 306 ThTyrAsnLeuAlaThrGlyValArgThrValIleValAlaGlySerGlnPheLeuIle 325
DB 2085 ATCTACCCGCTGGCGCGCGCATCCGCTCAGTGCATGAGGCGCAAGCTTCTGCATC 2144
QY 326 AsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGluAspThrAlaValArgIly 345
DB 2145 AACCAACAAACGTTCTACTTACTTACGCTTGGCCGCTCATGAAATGCGATTGGCGGCG 2204
QY 346 LysGlyHisAspProAlaTyrMetValHisAspPheGlnMetLysTrpIleGlyAla 365
DB 2205 AAAGGATTCATTAACGCTGCTGATGTGCACGATCAACGCTAATGACCTGGAATGGGGCC 2264
QY 366 AsnSerPheArgThrSerHisTyrProTyrAlaGluGluValMetAspPheAlaAspArg 385
DB 2265 AACTCTCACTACCTGCTGCTGCTTACCTTACGCTGAGAGATGCTGCATCTGGCAGATGAA 2324
QY 386 AsnGlyIleValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu---Met 404
DB 2325 CATGGCATCGTGGATGATGATGAAGCTGAGCTGTGGCTTTAAGCTCTCTTTAGGCAT 2384
QY 405 GlyValSerGluSerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLys 423
DB 2385 GGTTCGAAACGGCGCAACAAAGCGAAGAACTGTACAGCAAGAGCGAGTCAACGGGGAA 2444
QY 424 ThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAla 443
DB 2445 ACTCAGACGCGGCACTTACAGGCGATTAAAGACTGATACGCGCTGACAAACCAACCCA 2504
QY 444 SerValValMetTyrSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgIly 463
DB 2505 AGCGGTGTGATGTGAGTATTTGCAACGAAACCGGATCCGCTCGCAAGGTGACAGGAA 2564
QY 464 TyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPhe 483
DB 2565 TATTTGCGCGCACTGGCGGAGAAACCGCTAACTCATCCGACGCGCTCGATCACCCTGC 2624
QY 484 AlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSer 503
DB 2625 GTCAATGTATATGTTTCCGACGCTCACACCGATACCATCAGCATCTCTTATGATGCTG 2684
QY 504 CysIleAsnArgTyrPheGlyTyrTyrSerGlnThrGlyLysPheGluGluAlaGluAla 523
DB 2685 TGCCTGAACCGTTATTTACGGTTGATGTCMAAAGCCGCGATTTGGAAACGCAAGAG 2744
QY 524 AlaLeuGlnLysGluLeuHisGlyTyrGlnGluLysPheHisArgProIleValMetThr 543
DB 2745 GTACTGAAAAAGAACTTCTGGCTGCGAGAGAAACTGCATCAGCCGATATCATCACC 2804
QY 544 GlyTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyrSerGlu 563
DB 2805 GAATACGGCGCTGATACGTTACGCGGCTGCATCATGACACCGACATGCGAGTGA 2864
QY 564 GluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAspArgIleGluSerMet 583
DB 2865 GAGTATACAGTGTGATGATGCTGATGTATCATCCGCTCTTATGCTGCTCACGCGCTC 2924
QY 584 AlaGlyGlnHisValTyrAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgVal 603
DB 2925 GTGCGGTGAACAGATGAGAAATTTGCGCATTTTCCGACTTCCGAGGACATATTTGCGGTT 2984
QY 604 AspGlyAsnLysLysGlyValPheThrArgAspArgLysProLysAlaIleHisSer 623
DB 2985 GGGCGTACAAAGAAAGGAGATCTTACCCCGGACCCGAAACCGAAGTGGCGGCTTTTCTG 3044
QY 624 LeuArgAlaArgTyrThrSerIleAsp 632
DB 3045 CTGCAAAAAACGTGACTGCGATGAAAC 3071

```

```

XX 29-MAY-2003 (first entry)
XX DT
XX XX Plasmid pAgila nucleotide sequence SEQ ID NO:108.
DE
XX XX Chromosome-based platform; artificial chromosome; eukaryotic chromosome;
XX KM at site; integrase; recombinase; Aces; gene therapy; transgenic animal;
XX KM platform artificial chromosome expression system; gene; ds.
XX OS
XX Synthetic.
XX PN
XX MO200297059-A2.
XX PD
XX 05-DEC-2002.
XX PF
XX 30-MAY-2002; 2002MO-US017452.
XX PR
XX 30-MAY-2001; 2001US-0294758P.
XX PR
XX 21-MAR-2002; 2002US-0366891P.
XX XX
XX PA (CHRO-) CHROMOS MOLECULAR SYSTEMS INC.
XX PI Perkins E, Perez C, Lindenbaum M, Greene A, Leung J, Fleming E,
XX PI Stewart S, Shellard J;
XX DR WPI; 2003-140461/13.
XX XX
XX PT Novel eukaryotic chromosome comprising one or many att sites which
XX PT permits site-directed integration in the presence of lambda-integrase,
XX PT useful for site-specific recombination-directed integration of DNA of
XX PT interest.
XX PS
XX Example 12; Page 234-235; 272pp; English.
XX CC
XX The present invention describes a eukaryotic chromosome (I) comprising
XX CC one or several att sites, where an att site is heterologous to the
XX CC chromosome, and permits site-directed integration in the presence of
XX CC lambda-integrase. Also described: (1) a platform artificial chromosome
XX CC expression system (Aces) (II) comprising several att sites that participate
XX CC in recombinase catalyzed recombination; and (2) a method (M1) for
XX CC introducing a heterologous nucleic acid into a platform artificial
XX CC chromosome. (I) can be used in gene therapy. (M1) is useful for
XX CC introducing a heterologous nucleic acid molecule into a platform
XX CC artificial chromosome, preferably an Aces. (II) is useful for producing a
XX CC transgenic animal (e.g. a fish, insect, reptile, amphibian, arachnid, or
XX CC mammal) by introducing (II) by cell fusion, lipid-mediated transfection,
XX CC by a carrier system, microinjection, microcell fusion, electroporation,
XX CC microprojectile bombardment or direct DNA transfer into an embryonic
XX CC cell, preferably a stem cell or an embryo. (II) comprises a heterologous
XX CC nucleic acid that encodes a therapeutic product which is useful for
XX CC making a library of Aces comprising random portions of a genome. ACC44612
XX CC to ACC44732 and ABP96650 to ABP96657 represent sequences used in the
XX CC exemplification of the present invention
XX CC
XX SQ Sequence 3451 BP; 951 A; 707 C; 851 G; 942 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 6,15e-139 Length: 3451
XX Score: 1717.50 Matches: 348
XX Percent Similarity: 68.84% Conservative: 95
XX Best Local Similarity: 53.74% Mismatches: 175
XX Query Match: 51.21% Indels: 21
XX DB: Gaps: 11

```

```

US-10-757-093-4 (1-634) x ACC44711 (1-3451)
QY 13 AlaAlaProSerLeuGlyThrProAlaIleArgHisPheProArgAsnGlu----- 29
DB 1221 TCTGCACCGGATCTCAGATC-----GAATTCCTCCGCGCGCGGGAATTCAC 1268
QY 30 ---MetThrGlnHisGlnGlnProLeuIleLysValArgProGlnIleArgThrSerSerArg 48
DB 1269 TGGATCCCGCGGATCGATGCTTATG---TTACGTCCTGTAGAAACCCCAACCCGCT 1325

```

QY 49 GluLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu---AlaSerGlyLeuAsnAsp 67  
 DB 1326 GAATTAACAAAACCTCGAGCGCTGGGCAATCGTGGATCGGCAAACTGGCAATT 1385  
 QY 68 ThrAlaGlnProTrpThrAlaProLeuProLysGlyLeuGluCysProValProAlaSer 87  
 DB 1386 GAGCAGCGTTGGTGGAAAGCGGGTTACAGAAAGCCGGGCAATTGGTGGCCAGGAGT 1445  
 QY 88 TyrAsnAspIlePheIleSerArgGluIleHisAspHisValGlyTrpValTyrTrpGln 107  
 DB 1446 TTTCAGATCATGTCGCCGATGCAATATGCTGATATGTCGGCAACGCTGGTATCAG 1505  
 QY 108 ArgGluValIleValProLysGlyTyrPheSerGlnGluArgTrpLeuValArgAlaGluSer 127  
 DB 1506 CGCAAGCTCTTATACCGAAAGGTGGGCGAGCGGATACGCTGGCTTCGATGGC 1565  
 QY 128 AlaThrHisHisGlyArgIleTyrValAsnAsnArgLeuValAlaGluHisValGlyGly 147  
 DB 1566 GTCACTCATACGCAAAAGTGGGTCAATATCAGGAAGTGAATGAGCATACAGGGCGGC 1625  
 QY 148 TyrTrpProPheGluAlaAspValThrGluLeuValAlaProGlyGlyLysPheArgLeu 167  
 DB 1626 TATACCCCATTTGAAGCCGATGTCACCGCTATGTTATTCGCGGAAAGTGAACGATC 1685  
 QY 168 ThrIleGlyValAsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThr 187  
 DB 1686 ACGATTGTGTGAACCAAGCACTGAACCTGACAGCATATCCCGCGGAAAGTGAATACC 1745  
 QY 188 GlyAsnAlaThrGlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGly 207  
 DB 1746 ---GACGAAACCGCAAGAAAGCACTCTTACTTCATGATTTCTTAACTACGCCGG 1802  
 QY 208 LeuAlaArgSerIleTrpLeuTyrSerValProGlnGlnHisIleGlnAspIleThrVal 227  
 DB 1803 ATCCATCGACGCTAATGCTCTACACACCGCCGAAACCTGGGGAGCATATCACCGCTG 1862  
 QY 228 ValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGluValGluValAla 245  
 DB 1863 GTACCGCATGTCCGCAAGACTGAACCGCTGCTGTATCTGGCAGGTG---GTGGCC 1919  
 QY 246 AsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAla 265  
 DB 1920 AAT-----GCTATGTCAGCGTTGAACCTGGGTATGCGATCAACAGGTGTTGCA 1970  
 QY 266 LysAlaSerGlyAlaGlnGlyThrValThrIleProSerValLysLeuTrpGlnProGly 285  
 DB 1971 ACTGACAAAGCACACGCGGACCTTTCAGAGTGTGAATCCGCACTCTGGCAACCGGCT 2030  
 QY 286 AlaAlaIleTyrLeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValAlaAsp 305  
 DB 2031 GAAGTTATCTCATGAACCTGTACGTC-----ACAGCCAAAGCCGACAGAGTGTGAT 2084  
 QY 306 ThrTyrAsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIle 325  
 DB 2085 ATCTACCCGCTGGCGCGCTGCGCATCCGTCAGTGGCAGTGAAGGCGCAAGCTTCGATC 2144  
 QY 326 AsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGluAspThrAlaValArgGly 345  
 DB 2145 AACCAACAAACGCTTCTACTTACGCTTGGCGCTTCATGAAGATGGCGATTGGCGCGC 2204  
 QY 346 LysGlyHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTrpIleGlyAla 365  
 DB 2205 AAGAGATTCTCATTAACGCTGATGTGTGACAGATCAACCATTAATGAGCTGATGGGGCC 2264  
 QY 366 AsnSerPheArgThrSerHisTyrProTyrAlaGluGluValMetAspPheAlaAspArg 385  
 DB 2265 AACTCTCAACCGTACCTGCAATTAACCTTAAGCTGAAGATGCTCGACTGGCGCAGATGA 2324  
 QY 386 AsnGlyIleValValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu--Met 404  
 DB 2325 CATGGCATCGTGGTGAATGATGAACAGTCAAGCTGTCCGCTTAACTCTCTTACGATTT 2384

QY 405 GlyValSerGluSerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLys 423  
 DB 2385 GCTTTGAACCGCGGCAACAGCCGAAAGAACTGTACAGCAAGAGCATGCAAGGGGAA 2444  
 QY 424 ThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAla 443  
 DB 2445 ACTCACAGCGCGCATTAACGCGGATTAAGACTGATAGCCGCTGACAAACCAACCCCA 2504  
 QY 444 SerValValMetTrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGlu 463  
 DB 2505 AGCGTGTATGTGGAGTATGTCCAAAGCAACCGGATACCCGTCGCAAGGTGACGGGAA 2564  
 QY 464 TyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPhe 483  
 DB 2565 TATTTGCGCGCATTCGCGGAGCAACGCGTAACTGATCCGACCGCTCGCATCCTGC 2624  
 QY 484 AlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSer 503  
 DB 2625 GTCAATGTAAATGTTTCGCAAGCTCAACCGCATACATCAGCATCTTTGATGTGCTG 2684  
 QY 504 CysIleAsnAspArgTrpPheGlyTyrTrpTyrSerGlnThrGlyAspLeuGluAlaGluAla 523  
 DB 2685 TGCCTAACCCTTATTAACGTTGGTATGTCCAAAGCGCGCATTTGGAACGCAAGAG 2744  
 QY 524 AlaLeuGluLysGluLeuHisGlyTyrGlnGluLysPheHisArgProIleValMetThr 543  
 DB 2745 GTACTGAAAGAAACATTCCTGCGCTGCGAGAGAACTGATACCGCATATATCATCAC 2804  
 QY 544 GluTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGlu 563  
 DB 2805 GAATACGGCGGTGATACGTTAGCCGGGCTGCATCAATGATACCAACCATGTGAGTGA 2864  
 QY 564 GluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAspArgIleGluSerMet 583  
 DB 2865 GAGTATACAGTGCATAGCTGCGATATGTATCACCGGCTCTTATATGCGGCACCGCTC 2924  
 QY 584 AlaGlyGluHisValTrpAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgVal 603  
 DB 2925 GTCGGTGAACAGATATGAAATTTCCGCGATTTTGCACCTCGCAAGGCAATTCGCGGT 2984  
 QY 604 AspGlyAsnLysLysGlyValPheThrArgAspArgLysProLysAlaAlaHisSer 623  
 DB 2985 GCGGTATACAAAGAGGATCTTACCCGCGAACCGCAACCGAAGTGGCGGCTTTCTG 3044  
 QY 624 LeuArgAlaArgTrpThrSerIleAsp 632  
 DB 3045 CTGCAAAAACGCTGACGTGGCATGAAC 3071  
 DB  
 RESULT 4  
 ABT16610  
 ID ABT16610 standard; DNA; 14627 BP.  
 AC ABT16610;  
 XX  
 DT 03-APR-2003 (first entry)  
 XX  
 DE Artificial plant chromosome related plasmid DNA SEQ ID NO 21.  
 XX  
 KW Plant artificial chromosome; PAC; transgenic plant; vaccine;  
 KW blood factor; herbicide; stress; agronomical; nutrient quality;  
 KW bacterial artificial chromosome; BAC; yeast artificial chromosome; YAC;  
 KW de.  
 OS Unidentified.  
 OS  
 PN WO200296923-A1.  
 XX  
 PD 05-DEC-2002.  
 XX  
 PF 30-MAY-2002; 2002WO-US017451.  
 XX  
 PR 30-MAY-2001; 2001US-0294687P.  
 PR 04-JUN-2001; 2001US-0296329P.

XX (CHRO-) CHROMOSOME MOLECULAR SYSTEMS INC.  
PA (AGRI-) AGRISOMA INC.  
XX Perez C, Fabijanski SF, Perkins E;  
PI WPI; 2003-140436/13.  
XX  
PT Producing artificial chromosome by introducing a nucleic acid into plant  
PT cell, selecting artificial chromosome that has one or more repeat regions  
PT with equivalent amounts of euchromatic and heterochromatic nucleic acids.  
PS Example 5; Page 249-253; 269pp; English.  
XX  
XX The invention relates to a novel method for producing plant artificial  
XX chromosomes. The invention also relates to methods for targeting  
XX insertion of heterologous DNA into plant artificial chromosomes, methods  
XX for delivery of plant chromosomes to selected cells and tissues. The  
XX isolated plant artificial chromosome (PAC) is useful for producing a  
XX transgenic plant, which involves introducing the PAC into a plant cell.  
XX The PAC comprises a heterologous nucleic acid encoding a gene product  
XX such as enzymes, antisense RNA, tRNA, rDNA, structural proteins, marker  
XX proteins, ligands, receptors, ribozymes, therapeutic proteins, and  
XX biopharmaceutical proteins, vaccines, blood factors, antigens, hormones,  
XX cytokines, growth factors, antibodies, or a product that provides for  
XX resistance to diseases, insects, herbicides, or stress in a plant. The  
XX heterologous nucleic acid optionally encodes a product that provides an  
XX agronomically important trait in the plant, e.g. a product that alters  
XX nutrient use and/or improves the nutrient quality of the plant. The  
XX heterologous nucleic acid is contained within a bacterial artificial  
XX chromosome (BAC) or a yeast artificial chromosome (YAC). This  
XX polynucleotide sequence represents the DNA of a plasmid used in the  
XX method of the invention  
SQ Sequence 14627 BP; 3582 A; 3759 C; 3900 G; 3386 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 3.99e-138 Length: 14627  
Score: 1717.50 Matches: 338  
Percent Similarity: 68.84% Conservative: 95  
Best Local Similarity: 53.74% Mismatches: 175  
Query Match: 51.21% Indels: 21  
DB: Gaps: 11  
US-10-757-093-4 (1-634) x ABT16610 (1-14627)  
QY 13 AAlaAlaProSerLeuGlyThrProAlaAlaArgHisPheProArgAsnGlu----- 29  
DB 12069 TCTGCACCGGATCTCGAGATC-----GAATTCCCGCGCGCGCAATTCTAG 12116  
QY 30 --MetThrGlnHisGlnProLeuValArgProGlnArgThrSerSerArg 48  
DB 12117 TGGATCCCGCGGATCGGTACGTCCTTATG---TTACGTCTCTAGAAACCCACCGT 12173  
QY 49 GluLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu---AlaSerGlyLeuAsnAsp 67  
DB 12174 GAATCATAAAACTCGACGCGCTGTGGCATTCAGTCTGGATCGCAAACTGTGAGATT 12233  
QY 68 ThrAlaGlnProThrAlaProLeuProLysGlyLeuGluLysProValProAlaSer 87  
DB 12234 GAGCAGCGTGTGGGAAACCGCGTTACAGAAAGCCGGCAATTCTGTGCGAGCGAGT 12293  
QY 88 TyrAsnAspLeuPheLeuSerArgGluLeuHisAspHisValGlyTrpValTyrTrpGln 107  
DB 12294 TTTAACGATACAGTTCGCCGATGCAAGATATTCGTAATTATGTGGCAACGTCGTATCAG 12353  
QY 108 ArgGluValIleValProLysGlyTrpSerGlnGluArgTyrLeuValArgAlaGluSer 127  
DB 12354 CGGAGAGTCTTTATACGAAAGGTGGCGAGCGAGCGTATCGTGTGCTTCGATGCG 12413  
QY 128 AlaThrHisIleGlyArgIleTyrValAlaAsnAspArgLeuValAlaGluHisValGly 147  
DB 12414 GTCACTCATTAACGCAAAAGTGTGGTCAATAATACAGAAAGTATGATGAGATCAGGCGCGG 12473

QY 148 TyrThrProPheGlnAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeu 167  
DB 12474 TATACCCATTGTAACCCAGATGTCACCGCGTATGTTATTCGCGAAAGTACGATAC 12533  
QY 168 ThrIleGlyValAsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThr 187  
DB 12534 ACAGTTGTGTGAACAACGAACTGAACCTGGACAGACTATCCCGCGGAATGGTATAC 12593  
QY 188 GlyAsnAlaThrGlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGly 207  
DB 12594 --GACGAAACCGCAAGAAAGAGCTTACTTCATGATTTCTTAACTACACCGCGG 12650  
QY 208 LeuAlaArgSerIleThrLeuTyrSerValProGlnGlnHisIleGlnAspIleThrVal 227  
DB 12651 ATTCATCGACGCTAATGCTTACACACCGCAACCTGGGTGACATACCTACCGTG 12710  
QY 228 ValThrAspVal-----AspGlyAsnGlyLeuIleAsnTyrGluValGluValAla 245  
DB 12711 GTGACCATGTGGCGCAAGCTGTACACGCGTGTGTGATCGCAGGAGTG--GTGGCG 12767  
QY 246 AsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAla 265  
DB 12768 AAT-----GTTATGTCAAGCGTGAACCTCGATGCGGATCAGACAGGTGTGCA 12818  
QY 266 LysAlaSerGlyValaGlnGlyThrValThrIleProSerValLysLeuTrpGlnProGly 285  
DB 12819 ACTGGAACAGCACCGCGGACTTTGCAAGTGTGAATCCGCACTCTGCGACCGGCT 12878  
QY 286 AlaAlaTyrLeuTyrGlnLeuGlnValaIleValGlySerSerGlyAspValValaAsp 305  
DB 12879 GAAGGTATCTCATGAACCTGATACGTC-----ACAGCAAAAGCCAGACAGTGTGAT 12932  
QY 306 ThrTyrAsnLeuAlaThrGlyValaArgThrValLysValAlaGlySerGlnPheLeuIle 325  
DB 12933 ATCTACCCGCTGGCGTCCGATCGGATCGGCAAGTGAAGGCAACAGTTCCTGATC 12992  
QY 326 AsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGlyAspThrAlaValArgGly 345  
DB 12993 AACCAAAACCGTTCATCTTACTGCTTTGGCCGTCATGAAGATCGGATTTGCCCGG 13052  
QY 346 LysGlyHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTrpIleGlyAla 365  
DB 13053 AAAGGATGGAATACCTGCTGATGTCAGCATCAACGATTAATGATCGATTTGGGGCC 13112  
QY 366 AsnSerPheArgThrSerHisTyrProTyrAlaGluGluValMetAspPheAlaAspArg 385  
DB 13113 AACTCTACCGTACCTCGCATTAACCTTACGCTGAAGAGATGCTGACGTGGGAGATGAA 13172  
QY 386 AsnGlyIleValValIleAspGluThrProAlaValaGlyLeuAsnIleAlaLeu--Met 404  
DB 13173 CATGGATGTGTGTATGATGAACCTGACGCTGTGGCTTTAACTCTCTTATAGCATT 13232  
QY 405 GlyValSerGluSerGlyAlaPro--GlnThrPheThrProAspAlaIleAsnAspLys 423  
DB 13233 GGTTCGAAGCCGGCAACAGCGAAAGAACTGTACAGCGAAGGCACTCAACGGGGAA 13292  
QY 424 ThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAla 443  
DB 13293 ACTGACAGCGGACTTACAGCGATTAAGAGCTGAAGCGGTGACAAACCAACCCCA 13352  
QY 444 SerValValMetTrpSerIleAlaAsnGluProAlaSerHisGlyAspGlyAlaArgGlu 463  
DB 13353 AGCGTGTATGTGAGATTTTCCAAAGAACCGATACCCGTCGGAAGTGCACGGGAA 13412  
QY 464 TyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPhe 483  
DB 13413 TATTTTCGCCCACTGGCGGAACCAACGCTAAACTGATCCACGCGGTCCGATCACTTCG 13472  
QY 484 AlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSer 503  
DB 13473 GTCAATGTAATGTTCTGCGACCGCTCAACACGATACCATGAGATCTTTGATGTGCTG 13532

Qy	504	CyelleaahnaRgYrPhrPhegiLYTPRyYrSseGInThrGlyAAspleuGluuIaIaGluA	523
Db	13533	TGCGTGAACCGGTTATTACGGTTGGATGATGTCACAAAGCGCGATTGGAAACGGCAAGGAAG	13532
Qy	524	AlaleuGluIuIyegIuleuHiaGlyTYTPGInGluIuIyAphEhisArpProIleValMetThr	543
Db	13593	GTACTGGAAAAAGAACTTCTGGCGCTGGACGAGAAACCTGCATCAGCGGATTATCATCACC	13652
Qy	544	GIuTYrGlyAlaAAspThrLeuAlaGlyLeuHiaSerIleleuGlyLeuProTrpSerGlu	563
Db	13653	GAATATCGCGCGTGGATACGTATGACCGGGCTGCTCACTCATGTATACACCAATGTGGAGTGA	13712
Qy	564	GIuIphEgInValGIuInMetLeuAAspMetTYrHisArGValPheAAspArgIleGIuSerMet	583
Db	13713	GAGATATAGTGTGCATCGCTGGATATGTATACACCGGCTCTTGTATGCGGTACGCGCGTC	13772
Qy	584	AlaGIYGIuHiaValTYrPheAAspAlaAAspPheGIuThrAAspleuGIYIleIleAArgVal	603
Db	13773	GTCGGTGAACAGATATGGAATTTCCGCCGATTTTGGCACTCCGAAAGCATATTGCGCGTT	13832
Qy	604	AspGIYAsnLYsLYsGlyValAPhetHrArgAAspArgLYsProLYsAlaAlaHisSer	623
Db	13833	GGCGGTAAACAAGAAAGGGGATCTTCAACCCGCGAACCGAAACGAAAGTCGGCGCTTTCTG	13892
Qy	624	LeuAArgAlaArgTrpThrSerIleAsp	632
Db	13893	CTGCACAAAACGCTGGACTGGCATGCAAC	13919
RESULT 5			
ACC44712			
ID	ACC44712	standard; DNA; 14627 BP.	
XX	ACC44712;		
AC			
DT	29-MAY-2003	(first entry)	
XX			
DE	Plasmid pAG11a nucleotide sequence SEQ ID NO:109.		
XX			
KW	Chromosome-based platform; artificial chromosome; eukaryotic chromosome;		
KM	att site; integrase; recombinase; Aces; gene therapy; transgenic animal;		
XX	platform artificial chromosome expression system; gene; ds.		
OS	Synthetic.		
XX			
PN	MO200297059-A2.		
PN			
PD	05-DEC-2002.		
XX			
PF	30-MAY-2002; 2002MO-US017452.		
XX			
PR	30-MAY-2001; 2001US-0294758P.		
PR	21-MAR-2002; 2002US-0366891P.		
XX			
XX	(CHRO-) CHROMOS MOLECULAR SYSTEMS INC.		
XX			
PI	Perkins E, Perez C, Lindenbaum M, Greene A, Leung J, Fleming B,		
PI	Stewart S, Sheppard J;		
XX			
DR	WPI; 2003-140461/13.		
XX			
PT	Novel eukaryotic chromosome comprising one or many att sites which		
PT	permits site-directed integration in the presence of lambda-integrase,		
PT	useful for site-specific recombination-directed integration of DNA of		
XX	interest.		
XX			
PS	Example 12; Page 235-239; 272pp; English.		
XX			
CC	The present invention describes a eukaryotic chromosome (1) comprising		
CC	one or several att sites, where an att site is heterologous to the		
CC	chromosome, and permits site-directed integration in the presence of		
CC	lambda-integrase. Also described: (1) a platform artificial chromosome		
CC	expression system (Aces) (II) comprising several sites that participate		
CC	in recombinase catalysed recombination; and (2) a method (M1) for		

DB	Seq	Score	Length	Matches	Conservative	Mismatches	Indels	Gaps
CC	introducing a heterologous nucleic acid into a platform artificial chromosome. (I) can be used in gene therapy. (M) is useful for introducing a heterologous nucleic acid molecule into a platform artificial chromosome, preferably an Acell. (II) is useful for producing a transgenic animal (e.g. a fish, insect, reptile, amphibian, arachnid, or mammal) by introducing (II) by cell fusion, lipid-mediated transfection by a carrier system, microinjection, microcell fusion, electroporation, microprojectile bombardment or direct DNA transfer into an embryonic cell, preferably a stem cell or an embryo. (II) comprises a heterologous nucleic acid that encodes a therapeutic product which is useful for making a library of Acells comprising random portions of a genome. ACC44612 to ACC44732 and AB996550 to AB996577 represent sequences used in the exemplification of the present invention	3.99e-138	14627	38	95	175	21	11
US	10-757-093-4 (1-634) x ACC44712 (1-14627)	1717.50	14627	38	95	175	21	11
OY	13 AAlaIARSeRleuGlyThrProAlaAlaArgHisPheProArgHisGlu-----	12116	12117	12117	12117	12117	12117	12117
DB	12069 TCTGCACCGGATCTCGAGATC-----GAATTCCTCCGCGCGCAATTCACTAG	12116	12117	12117	12117	12117	12117	12117
OY	30 ---MetThrGlnHisGluGlnProLeuIleIysValArgProGlnArgHisSerArg	12173	12174	12174	12174	12174	12174	12174
DB	12117 TGGATCCCGCGGATACGTCATCCTTTAG---TTACGTCCTGTAGAAACCCCAACCGT	12173	12174	12174	12174	12174	12174	12174
OY	49 GluIleuValAsnIleuAspGlyLeuTrpIysPheAlaIeu---AlaSerGlyLeuAsnSer	12233	12234	12234	12234	12234	12234	12234
DB	12174 GAATCAAAAAATTCAGACGCGCTGTGGCATTCAGTCGTGATCGCAAAACGTGGAAT	12233	12234	12234	12234	12234	12234	12234
OY	68 ThrAlaGlnProTrpThrAlaProLeuProIysGlyLeuGlnCybProValProAlaSer	12293	12294	12294	12294	12294	12294	12294
DB	12234 GAGCAGCGTGTGGGAAAGCGGCTTACAAAGACCGGCAATTGCTGTCCAGGCACT	12293	12294	12294	12294	12294	12294	12294
OY	88 TyrAsnAspIlePheHisLeuSerArgGluIleHisAspHisValGlyTrpValTyrTrpGln	12353	12354	12354	12354	12354	12354	12354
DB	12294 TTTAACGATCAGTTCGCCGATGCAAGATATTCGTAATTAATGTGGGCAACGCTGGTATCAG	12353	12354	12354	12354	12354	12354	12354
OY	108 ArgGluValIleValProIysGlyTrpSerGlnGluArgTyrLeuValArgAlaGluSer	12413	12414	12414	12414	12414	12414	12414
DB	12354 CGGAAAGCTTTTATACCGAAAGGTGTGGCAGCGCCAGCCAGATAGTGTGCTGTTCATGG	12413	12414	12414	12414	12414	12414	12414
OY	128 AlaThrHisHisGlyArgIleTyrValAsnAsnArgLeuValAlaGlnHisValGlyGly	12473	12474	12474	12474	12474	12474	12474
DB	12414 GTCACTCATTAAGCGCAAAAGTGTGGCTCAATTAATCAGAAAGTGTGAGCATCAGGGCGGC	12473	12474	12474	12474	12474	12474	12474
OY	148 TyrThrProPheGluAlaAspValIleArgIleuValAlaPheGlyGlyIysPheArgLeu	12533	12534	12534	12534	12534	12534	12534
DB	12474 TATAACGCATTGTAAACCCGATGTCAAGCCGATATGTATTATCCGCGGCAAAAGTCAATATAC	12533	12534	12534	12534	12534	12534	12534
OY	168 ThrIleIleValAlaAsnGlnGluLeuThrHisGluThrIlePheProGlyIysIleThrThr	12593	12594	12594	12594	12594	12594	12594
DB	12534 ACGATTGTGTGTGAACAACGAACTGGAACCTGGCAGACTATCTCCGCGGGAATGGGATTAC	12593	12594	12594	12594	12594	12594	12594
OY	188 GlyAsnAlaThrGlyAsnArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGly	12650	12651	12651	12651	12651	12651	12651
DB	12594 ---GACGAAAGCGCAAGAAAGAAAGCAAGCTTATCTTCAAGATTTCTTAACTACGCGGG	12650	12651	12651	12651	12651	12651	12651
OY	208 LeuAlaArgSerIleTrpLeuTyrSerValProGlnGlnHisIleGlnAspIleThrVal	12710	12711	12711	12711	12711	12711	12711
DB	12651 ATCATCGCAGCCTAATGTCTTACACACGCGCAACCACTGGGTGGACATATCACCGGTG	12710	12711	12711	12711	12711	12711	12711
OY	228 ValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnThrGlyValGluValAla	12767	12768	12768	12768	12768	12768	12768
DB	12711 GTAGCGATATTCGCGCAAGCTGTAAACACGCGCTCTGTGTACTGTGGAGGTGG---GTGGCC	12767	12768	12768	12768	12768	12768	12768

```

QY 246 AsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAla 265
Db 12768 AAT-----GGTGAATGTCAGCGCTTAACCTGGCTGATGGGATCAACAGCTGTTGCA 12818
QY 266 LysAlaSerGlyAlaGlnGlyThrValThrIleProSerValIlySLeuTPGInProGly 285
Db 12819 ACTGACACAGACAGACAGCGGAGCTTTGGCAATGGTGAATCCGCACTTGGCAACCGGGT 12878
QY 286 AlaAlaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 305
Db 12879 GAAGCTTATCTCTATGAACTTACGTC-----ACAGCCAAACCCAGACAGAGTGTGAT 12932
QY 306 ThrTyraSerLeuAlaThrGlyValArgThrValIlySLeuIleuIleuIleuIleuIleu 325
Db 12933 ATCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12992
QY 326 AsnGlyIlyProPheIlyPheThrGlyPheGlyIlySLeuAspThrAlaValArgGly 345
Db 12993 AACCAACAAACCGCTCTACTTACGCTTGGCGCTGATGAGATGGCGATTTGGCGCGC 13052
QY 346 LysGlyIlyAspProAlaIlyMetValIlyAspPheIleuMetIlySTPILieGlyAla 365
Db 13053 AAAGGATTCGATTAACGTCGATGATGACGATCAACGATTAATGACCTGATTTGGGCGC 13112
QY 366 AsnSerPheArgThrSerHisIlyTyraIleGluIleuIleuIleuIleuIleuIleuIleu 385
Db 13113 AACTCTACCGCTACCTGCAATTAACCTTACGCGTAAGAGATGCTGCGATGGCGAGTGA 13172
QY 386 AsnGlyIleValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 404
Db 13173 CATGGCATCGCTGGATGATGAACTGACGCTGCTGCTTAACTCTCTTTAGGCGAT 13232
QY 405 GlyValSerGluSerGlyAlaPro---GlnThrPheThrProAspAlaIleAspIlyS 423
Db 13233 GGTTCGAAAGCGGCAACAGCCGAAAGACTGACGAGGAGGAGGAGGAGGAGGAGGAGGAG 13292
QY 424 ThrGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 443
Db 13293 ACTGACAGCGCGCTTACAGCGGATTAAGAGCTGATGACGCGTGAACAAACCAACCA 13352
QY 444 SerValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 463
Db 13353 AGCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 13412
QY 464 TyrPheGluProLeuThrAsnLeuThrArgGlnIleuAspProThrArgProIleuIleu 483
Db 13413 TATTTCCGCGCACTGGCGGAGAGCAAGCGGTAACCTGATCCGATCCGATCCCTGC 13472
QY 484 AlaAsnValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 503
Db 13473 GTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 13532
QY 504 CysIleAsnArgIlyPheGlyIlyPheIlyPheIlyPheIlyPheIlyPheIlyPheIlyPhe 523
Db 13533 TGGCTGAACCGTATTAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 13592
QY 524 AlaLeuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 543
Db 13593 GTACGAAAAAGAACTTGGCGCTGCGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 13652
QY 544 GlnIlyGlyAlaAspThrIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 563
Db 13653 GAATACCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 13712
QY 564 GluPheGlnValGlnMetLeuAspMetIlyHisArgValIleAspArgIleGluSerMet 583
Db 13713 GAATATCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 13772
QY 584 AlaGlyIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 603
Db 13773 GTGGGTGAACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 13832

```

```

QY 604 AspGlyAsnIlySLeuGlyValPheThrArgAspArgIlyProIlyValAlaIleuIleuSer 623
Db 13833 GCGCGTACAAAGAGGAGATCTTACCCCGGACCCGAAACCGAAGTGGCGCTTTCTG 13892
QY 624 LeuArgAlaArgTrpThrSerIleAsp 632
Db 13893 CTGCAAAAACGCTGACTGCGATGAAC 13919

RESULT 6
ID AAA27980 standard; DNA; 4299 BP.
XX
AC AAA27980;
XX
DT 15-AUG-2000 (first entry)
XX
DE Vector MS23-pBT10-GUS nucleotide sequence.
XX
KM Elicitor-responsive cis-element; vector MS23-pBT10-GUS; herbicide; ss;
KM chimeric promoter; pathogen infection; transgenic plant; resistance;
KM local response; genetic engineering; disease resistant crop.
XX
OS Synthetic.
XX
PN WO200029592-A2.
XX
PD 25-MAY-2000.
XX
PF 12-NOV-1999; 99NO-EP008710.
XX
PR 12-NOV-1998; 98BP-00121160.
PR 27-AUG-1999; 99BP-00116981.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG.
XX
PT Kirsch C, Logemann E, Hahlbrock K, Rushton P, Somseich I,
PT WPI; 2000-387804/33.
XX
DR Chimeric promoters mediating gene expression in plants upon pathogen
PT infection, useful for transgenic plant production comprises at least one
PT cis-acting element for direct elicitor-specific expression.
XX
PS Example; Fig 1; 73bp; English.
XX
CC This sequence represents a synthetic vector MS23-pBT10-GUS encoding
CC nucleotide sequence. The present invention relates to chimeric promoters
CC capable of mediating local gene expression in plants upon pathogen
CC infection. The chimeric promoters comprise at least one cis-element (see
CC AAA27964-A27979) capable of directing elicitor-specific expression, and a
CC minimal promoter. The chimeric promoters are useful for producing a
CC transgenic plant which has attained resistance or improved resistance
CC against a pathogen. The cis-acting element, chimeric promoter,
CC recombinant gene encoding the chimeric promoter, vector comprising the
CC chimeric promoter and a compound capable of activating the chimeric
CC promoter are useful for producing pathogen resistant plants, and for
CC identifying and/or producing compounds capable of conferring induced
CC resistance to a pathogen in a plant. A compound which specifically
CC activates or inhibits genes activated in a plant when attacked by a
CC pathogen is also useful as a plant protective agent or a herbicide. The
CC chimeric promoter provides rapid and local response to pathogen attack
CC but shows negligible activity in uninfected parts of the plants and
CC therefore is most suitable for the engineering of disease resistant crops
XX
SO Sequence 4299 BP; 1107 A; 1028 C; 1123 G; 1041 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 9,03e-139 Length: 4299
Score: 1717.00 Matches: 332
Percent Similarity: 70.02% Conservative: 93
Best Local Similarity: 54.70% Mismatches: 166
Query Match: 51.19% Indels: 16
Gaps: 9

```



```

US-10-757-093-4 (1-634) x AAA27980 (1-4299)
QY 33 HIGLGLINProLeuIlelyValArgProGLNArgThiSerSerArgIuLeuValAsn 52
DB 2103 CAGCGCTCGAGTGGCCACCATGATCCGCTCGTAGMAAACCACCCCTGAAATGAAAAA 2162
QY 53 LeuAspGlyLeuThrPlyPheAlaLeu-----AlaSerGlyLeuAnaArgThrAla 69
DB 2163 CTGACGCGCTGTGGCATTCAGTCTGGATCCGCAAAACTGTGGAATGATCAG----- 2216
QY 70 GlnProThrAlaProLeuProlyGlyLeuGluCySProValProAlaSerTyAsn 89
DB 2217 CGTTGGGGGAAAGCGCGCTTACAAGAAAGCCGCAATGCTGTGCCAGGCAAGTTTAAAC 2276
QY 90 AspIlePheIleSerArgIuIleHisAspHisValIleTyTrValIleTyTrGlnArgGlu 109
DB 2277 GATCAGTTCCCGCATGACGATATTCGTAAATATGTCGGGCAACGCTCGTATCAGCCGCA 2336
QY 110 ValIleValProlyGlyTrpSerGlnIuArgTyTrLeuValArgAlaGluSerAlaThr 129
DB 2337 GTCTTATACGAAAGGTGGGAGCCAGCCAGGTATCGTGGCTTTCGATGCGGTACT 2396
QY 130 HisHisGlyArgIleTyTrValAsnAsnArgLeuValAlaGluHisValIleGlyTyTrThr 149
DB 2397 CATTAACGCAAAAGTGGGTCAATATCAGGAAGTGAAGATCAGGCGCGGTATACG 2456
QY 150 ProPheGluAlaAspValIleThrGlyLeuValAlaProGlyGluTyPheArgLeuThrIle 169
DB 2457 CCATTTGAAACCGCATGTACAGCCGCTATGTTCCCGGAAAGTGTACCTATCACCGCTT 2516
QY 170 GlyValAsnAsnGlnLeuThrHisGluThrIleProProGlyTyIleThrThrGlyAsn 189
DB 2517 TGTGTGAACAACGAACTGAACCTGGACAGATATCCCGCGGAATGGATGATTACC--GAC 2573
QY 190 AlaThrGlyIleAspArgIleGlnThrTyTrGlnHisAspPheTyTrAsnTyTrAlaGlyLeuAla 209
DB 2574 GAAACGCGCAAGAAAGACAGTCTTACTTCCATATTTCTTTAACTTATGCCGGAATCAT 2633
QY 210 ArgSerIleTrpLeuTySerValProGlnGlnHisIleGlnAspIleThrValIleThr 229
DB 2634 CGCAGCCTAATGCTTACACACCCCGAACACCTGGTGGACGATATCACCGTGTGACG 2693
QY 230 AspVal-----AspGlyAspAsnGlyLeuIleAsnTyTrGluValAlaIleAsnGln 247
DB 2694 CATGTCCGCAAGACTGAACACGCGCTGTGTGACGTGCAAGG--GTGGCCAAAT-- 2747
QY 248 ThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAlaIleValAla 267
DB 2748 -----GTTGATGTCAAGCTTGAACCTGCGATCCGATCAACAGTGTGCAACTGCA 2801
QY 268 SerGlyAlaGlnGlyThrValIleThrIleProSerValIleAspLeuTrpGlnProGlyAlaAla 287
DB 2802 CAAGGCACTAGCGGCACTTGGCAAGTGTGAATCCGCACTTGGCAACCGGCGTGAAGT 2861
QY 288 TyTrLeuTyTrGlnLeuGlnValIleAsnIleValIleGlySerSerGlyAspValIleAspThrTy 307
DB 2862 TATCTCTATGAACGTGCGTC-----ACAGCCAAAGCCAGACAGAGTGTGATATCTAC 2915
QY 308 AsnLeuAlaThrGlyValArgThrValIleValAlaIleGlySerGlnPheLeuIleAsnGly 327
DB 2916 CCGCTTCGCGCGCATCCGCTGCTGAGTGAAGGCGCAAGCTTCTCGATTAAACAC 2975
QY 328 LysProPheTyTrPheThrGlyPheGlyLysHisGluAspThrAlaValArgIleLysGly 347
DB 2976 AAACCGTTTACTTACTTGTGCTTGTGTCATGATGAAATGCGGCACTGCGCGCAAGGA 3035
QY 348 HisAspProAlaTyTrMetValHisAspPheGlnLeuMetLysTrpIleGlyAlaAsnSer 367
DB 3036 TTGCATTAACGTCTGATGTGACAGCACCATTAATGAATGATGGGCGCAACTCC 3095
QY 368 PheArgThrSerHisTyTrProTyTrAlaGluIleValMetAspPheAlaAspArgAsnGly 387
DB 368 PheArgThrSerHisTyTrProTyTrAlaGluIleValMetAspPheAlaAspArgAsnGly 387

```

```

DB 3096 TACCGTAACTCGCATTAACCTTACGCTGAAGAGATGCTCGACTGGCGAGATGAACATGCG 3155
QY 388 IleValValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu---MetGlyVal 406
DB 3156 ATCGGTGATGATGATAAACTGCTGCTGGCTTTAACTCTTTAGGCATTGGTTTC 3215
QY 407 SerGluSerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLysGln 425
DB 3216 GAAGCGGCAACAAGCCGAAAGAAAGTGTACAGCGAAGAGCATCAACGGGAAACTCAG 3275
QY 426 GlnAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAlaSerVal 445
DB 3276 CAAGCCGACTTACAGCGATTAAGAAGCTGATGCCGCTGACAAAACCCACCAAGCGTG 3335
QY 446 ValMetTrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgIuTyTrPhe 465
DB 3336 GTGATGTGGAGTGTGCCAAGCAAGCATCCGCTCCGCAAGTGTACAGGAATATTTTC 3395
QY 466 GluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsn 485
DB 3396 GCGCCACTGGCGGAGCAACGCTTAACTGACCCGACGCTCCGATCAGTCCGTCAT 3455
QY 486 ValGlyThrAlaThrTyTrGlnLeuAspArgIleSerAspLeuPheAspValSerCySile 505
DB 3456 GTATGTTTGGACGCTCACACCGATACCATCAGCATCTTTGATGTGCTGCTGCTG 3515
QY 506 AsnArgTyTrPheGlyTyTrPlySerGlnThrGlyAspLeuGluIuAlaIleAlaLeu 525
DB 3516 AACGTTATTAACGATGTGATGTCAAAGCGGGAATTTGGAAACGCGACAGAAAGTACTG 3575
QY 526 GluLysGluLeuHisGlyTrpGlnGluLysPheHisArgProIleValMetThrGluTyTr 545
DB 3576 GAAAGAACTTGTGGCTGCGCAGAGAAATCTGATCAGCCATTTATCATACCGAATAC 3635
QY 546 GlyAlaAspThrLeuAlaGlyLeuHisAspIleLeuGlyLeuProTrpSerGluIuPhe 565
DB 3636 GCGGTGATACGTTAGCGCGGCTGCACTCAATGATACACCAATGTGAGATGAAGATAT 3695
QY 566 GlnValGlnMetLeuAspMetTyTrHisArgValPheAspArgIleGluSerMetAlaGly 585
DB 3696 CAGTGTGCATGCTGTGATGTATCAACCGCTTGTGATGCTGTGACGCGCTGTGCTGT 3755
QY 586 GlnHisValIleTyTrAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGly 605
DB 3756 GAAACAGTATGAATTTTCCGCAATTTTGGCACTTCGCAAGCATATTTGCGGTGGCGGT 3815
QY 606 AsnLysLysGlyValPheThrArgAspArgLysProLysAlaAlaHisSerLeuArg 625
DB 3816 AACCAAGAAAGGATCTTCACTCGCAACCGCAACCGAAGTGGCGGCTTTTCTGTGCA 3875
QY 626 AlaArgTrpThrSerIleAsp 632
DB 3876 AAACGCTGACCTGCATGAAC 3896
RESULT 7
AAT38397 standard; DNA: 3169 BP.
AC AAT38397;
XX
XX
XX 25-MAR-2003 (revised)
XX 14-FEB-1997 (first entry)
XX
XX pTTC 99 plasmid fragment encoding recombinant antibody.
XX Antibody; fusion protein; recombinant antibody; tumour therapy; produg;
XX ss.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX FT 3..644
XX CDS /*tag= a
FT

```



```

Qy 472 ThrArgGlnLeuAspProThrArgProIleThrPheAlaIleValGlyThrAlaThrTyr 491
Db 2646 ACCGGTAACCTCGAACCGGACCGCTCGATACCTGCTCAATGTAATGTTCTGGACGCT 2705
Qy 492 GlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTyr 511
Db 2706 CACACCGATACCATACGCGATCTCTTGAATGTGTGTCGCGAACCCTTATACGGAATGG 2765
Qy 512 TyrSerGlnThrGlyAspLeuGluAlaGluAlaIleValGlyGluLeuHisGly 531
Db 2766 TATGTCGAACCGCGCGATTGGAAACCGCAGAGAAAGGTACTGGAAAAAGAACTTCTGGCC 2825
Qy 532 TrpGlnGluPhePheHisArgProIleValMetThrGlnTyrGlyAlaAspThrLeuAla 551
Db 2826 TGGCAGAGAAATGCACTACGCGATATATCATCCGAAATACGCGCTGGATACGTTAGCC 2885
Qy 552 GlyLeuHisSerIleLeuGlyLeuProTyrSerGluGluPheGlnValGlnMetLeuAsp 571
Db 2886 GGGCTGCACCTCATGTACACCGACATGTGAGTGAATACATCATGTGATGCGATGCGAT 2945
Qy 572 MetTyrHisArgValPheAspArgIleGluSerMetAlaGlyGluHisValTTPAspPhe 591
Db 2946 ATGTATACACCGCGCTTGTGATCGCGTCAGCGCGCTCGCTGCAACAGTATGGAATTTC 3005
Qy 592 AlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnValGlyValAlaPhe 611
Db 3006 GCCGATTTTCGACCTTCGCAAGGCAATATGCGCGCTTCGCGGTACCAAGAAAGGATCTTC 3065
Qy 612 ThrArgAspArgIlyProLysAlaAlaAlaHisSerLeuArgAlaArgTyrPheSerIle 631
Db 3066 ACTCGGACCGCAACCGAAGTGGCGGCTTTCTCTGCAAAAACGCTGACCTGCGCATG 3125
Qy 632 Asp 632
Db 3126 AAC 3128

RESULT 8
AAF80288
ID AAF80288 standard; DNA; 8654 BP.
XX
XX AAF80288;
XX
XX 29-JUN-2001 (first entry)
DE Nucleotide sequence of plasmid pMR1196.
XX
XX Vector; transgenesis; trfA locus; RK2 ori; oriV; P285 protein;
KM P382 protein; antibiotic resistance gene; nptII; transgenic plant; ss.
XX
XX Synthetic.
OS
FH Key 1. Location/Qualifiers
FH rep_origin 1.654
FT /tag= a
FT /note= "ori RK2"
FT 655.1263
FT /tag= b
FT /note= "ori COLB1"
FT 1264..2603
FT /tag= c
FT /note= "NPT III gene coding for neomycin
FT phosphotransferase and kanamycin resistance"
FT 2604..4098
FT /tag= d
FT /note= "trfA locus from RK2 coding for two proteins P285
FT and P382 enabling an increase in the replication rate"
FT 4106..4271
FT /tag= e
FT /note= "ori DNA left border"
FT 4272..4559
FT /tag= f
FT /note= "nopaline synthetase terminator"
FT

```

```

FT CDS 4560..5556
FT /tag= g
FT /note= "NPT III gene coding for neomycin
FT phosphotransferase and kanamycin resistance"
FT 5557..5771
FT /tag= h
FT /note= "nopaline synthetase promoter"
FT 5818..7717
FT /tag= i
FT /note= "GUS gene coding for beta glucuronidase"
FT 7718..8447
FT /tag= j
FT /note= "polyA from 35S ribosome"
FT 8474..8647
FT /tag= k
FT /note= "ori DNA right border"
XX
XX PR2798139-A1.
XX
XX 09-MAR-2001.
XX
XX 03-SEP-1999; 99FR-00011112.
XX
XX 03-SEP-1999; 99FR-00011112.
XX
XX (MER1-) MERISTEM THERAPEUTICS SA.
XX
XX Gruber V, Comeau D;
XX
XX WPI; 2001-259847/27.
XX
XX New vector free from non-essential elements, useful for transforming
XX cells for protein production and for preparing transgenic plants.
XX
XX Claim 20; Page 132-135; 180pp; French.
XX
XX The specification describes a synthetic vector containing only those
XX elements essential for its functionality and transgenesis of a cell
XX (especially a plant cell). The vector consists of at most one origin of
XX replication (ori), at most one sequence encoding a selection agent and a
XX trfA locus encoding a protein that increases the level of plasmid
XX replication. The vector particularly contains an RK2 ori, especially oriV
XX from PRK2 of Escherichia coli with a broad host range, an antibiotic
XX resistance gene (especially nptII conferring resistance to kanamycin in
XX bacteria) and a trfA locus from PRK2 encoding the proteins P285 and P382.
XX The vectors are used to prepare transgenic plants and transformed host
XX cells for production of a heterologous proteins, e.g. insulin,
XX interferon, lipase, blood proteins and anti-inflammatory agents. The
XX present sequence represents a plasmid of the invention
XX
XX Sequence 8654 BP; 2243 A; 2134 C; 2352 G; 1925 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No. 3.68e-138 Length: 8654
Score: 1714.50 Matches: 337
Percent Similarity: 69.12% Conservative: 95
Best Local Similarity: 53.92% Mismatches: 177
Query Match: 51.12% Indels: 17
DB: 4 Gaps: 9
US-10-757-093-4 (1-634) x AAF80288 (1-8654)
Qy 15 ProSerLeuGlyThrProAlaAlaAlaGHisPheProArgAsnGluMetThrGlnHisGlu 34
Db 5742 CCGAAGCTTGGCGCGCGGTAAACCGGTGATCTT--AAATTAATCGACCTTAGGGGT 5799
Qy 35 GlnProLeuIleLysValArgProGlnArgThrSerSerArgGluLeuValaIleAsp 54
Db 5800 GGTCAAGTCCCTTATGTTAGCTCTCTGTGAAGAACCCCAACCGGTAAATCAAAAACTCGAC 5859
Qy 55 GlyLeuTrpLysPheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnPro 71
Db 5860 GGCCTGGGCAATTCAGTGTGATCGCAAAACCTGGAATGATCAG-----CGTTGG 5913

```

QY 72 TTPThrAlaProLeuProLysGlyLeuGlucyAspProValProAlaSerTyrAsnAspIle 91  
 Db 5914 TGGGAAGCCGGCTTACAGAAAGCCGGCAATTCCTGTCGCGAGCGCTTTTACAGATCAG 5973  
 QY 92 PheIleSerArgGluIleHisAspHisValGlyTyrValTyrTyrGlnArgGluValIle 111  
 Db 5974 TTGGCCGATGCAGATATTCGTAATTCGCGGAAGCTGTGATATCAGCGCGAAGTCTTT 6033  
 QY 112 ValProLysGlyTyrPheSerGlnGluArgTyrLeuValArgAlaGluSerAlaThrHisHis 131  
 Db 6034 ATACCGAAAGGTTGGGAGCGCGAGCGATATCGTGTGCTGCTTCATGCGGTGCTACCTATAC 6093  
 QY 132 GlyArgIleTyrValAsnAsnArgLeuValAlaGluHisValGlyGlyTyrThrProPhe 151  
 Db 6094 GCGAAGTGGCGGTCAATATCAGAGATGATGAGATCAGAGCGCGCTATACGCCATTT 6153  
 QY 152 GluAlaAspValThrGluLeuValAlaProGlyGlyLeuPheArgLeuThrIleGlyVal 171  
 Db 6154 GAAGCCGATGTACGCGGTATGTTATGCGCGGAAAGTGTATGATACCGTATGCTGTG 6213  
 QY 172 AsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThr 191  
 Db 6214 AACGACGAACTGAACTGCGACGACTATCCCGCGGAATGATGATACC--GACGAAAC 6270  
 QY 192 GlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSer 211  
 Db 6271 GCGAAGAAAGAGCGTTTACTTCCATGATTTCTTTTAACTATGCGGAATCCATCCGACG 6330  
 QY 212 IleTyrLeuTyrSerValProGlnGlnHisIleGlnAspIleThrValIleThrAspVal 231  
 Db 6331 GTATGCTCTTACACCGCGCAACACCTGGTGACATATCACCGGTGGAGCGCATGTC 6390  
 QY 232 -----AspGlyAspAsnGlyLeuIleAsnTyrGluValGluValAlaAsnGlnThrThr 249  
 Db 6391 GCCCAAGACTGTAAACCGCGCTCTGTGACTGGAGGTG--GTGGCAAT----- 6438  
 QY 250 GlyGlnIleGlnIleSerValIleAspGlyAspGlyAlaIleValAlaLysAlaSerGly 269  
 Db 6439 GGTGATGTACGCGTTGAATCGCGTGAATGCGGATCAACAGGTGTTGAACTGACAAAGC 6498  
 QY 270 AlaGlnGlyThrValThrIleProSerValLysLeuTyrGlnProGlyAlaIleTyrLeu 289  
 Db 6499 ACTAGCGGAGCTTGCAGAGTGAATCCGACCTCTGGCAACCGGCTGAAGGTATATCTC 6558  
 QY 290 TyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValValAspThrTyrAsnLeu 309  
 Db 6559 TATGAATGTGCGTC-----ACAGCCAAAGCCAGACAGATGTGATATCTACCGCTT 6612  
 QY 310 AlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysPro 329  
 Db 6613 CGCGTCGCGCATCCGTCAGTGCAGTGAAGGCGGAACAGTTCCTGATTAACCAAAACG 6672  
 QY 330 PheTyrPheThrGlyPheGlyLysHisGluAspThrAlaValArgGlyLysGlyHisAsp 349  
 Db 6673 TTCTACTTTTACTGCTTGTGCTGCTTGAAGATCGGACGATGTTGGTGGCAAAAGATTCGAT 6732  
 QY 350 ProAlaTyrMetValHisAspPheGlnLeuMetLysTyrIleGlyAlaAsnSerPheArg 369  
 Db 6733 AACGTGCTGATGTGACAGCAACGATTAATGATGATGATGATGATGATGATGATGATGAT 6792  
 QY 370 ThrSerHisTyrProTyrAlaGluGluValAlaMetAspPheAlaAspArgGlyIleVal 389  
 Db 6793 ACCTCCCATTAACCTTAACGCTGAAGATGCTGATGCGGCAATGAACATGCGATCGTG 6852  
 QY 390 ValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeu--MetGlyValSerGlu 408  
 Db 6853 GTGATGATGAAGAACTGCTGCTGCTGCTTAACTCTCTTTAGCATGTTTCGAAAGCG 6912  
 QY 409 SerGlyAlaPro--GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAla 427  
 Db 6913 GGCACACAGCCGAAAGAACTGTACAGCGAAGAGCGAGTCAACGCGGAACCTACGCAAGCG 6972

QY 428 HisLeuGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAlaSerValIleMet 447  
 Db 6973 CACTTACAGCGCATTTAAAGACTGATAGCGCGTGAAGAAACCAACCAAGCGGTGATG 7032  
 QY 448 TrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGlyLysThrPheGluPro 467  
 Db 7033 TGGAGTATGCCAACCAACCGGATACCCGTCGCAAGGTGACAGCGGAATATTTCCGCCCA 7092  
 QY 468 LeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGly 487  
 Db 7093 CTGGCGAAGCAACGCGTAACTCGAACCGCGCTCATGATACCTGCTCATATGTAATG 7152  
 QY 488 ThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArg 507  
 Db 7153 TTCTGGAGCGCTCACACCGATACATCACCGATCTTTGATGTGTGTGCTGCCAAGCCT 7212  
 QY 508 TyrPheGlyTyrTyrSerGlnThrGlyAspLeuGluGluAlaAlaLeuGluLys 527  
 Db 7213 TATTACGATGGTATGTCCAAAGCGCGATTTGAAACCGCGAGAGAGTACTGGAAAAA 7272  
 QY 528 TyrPheGlyTyrGlnGluLysPheHisArgProIleValMetThrGluTyrGlyAla 547  
 Db 7273 GAACCTTCCGCTGCGAGAGAACTGCATCGACCGATTAATCATCCGAAATACGCGCTG 7332  
 QY 548 AspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyrSerGlnGluThrVal 567  
 Db 7333 GATACCTTACCGCGGTGACATCAATGATACCGCATGTGGAGTGAAGATATCATCGTGT 7392  
 QY 568 GlnMetLeuAspMetTyrHisArgValPheAspArgIleGluSerMetAlaGlyLysHis 587  
 Db 7393 GCATGCTGTGATATGATATACCGCGCTTGTGATGCGGTGACGCGCTCTGCGTGAACG 7452  
 QY 588 ValTyrAsnPheAlaAspPheGlnThrAsnLeuGluIleIleArgValAspGlyLeuLys 607  
 Db 7453 GTATGGAATTTGCCCATTTTGGACCTGCGAAGGATATATGCGGTGCGGTAAACAG 7512  
 QY 608 LysGlyValPheThrArgAspArgLysProLysAlaAlaAlaHisSerLeuArgAlaArg 627  
 Db 7513 AAAGGATCTTCACTGCGACCGCAACCGGAAGTGGCGGCTTTCTGCTGCAAAAAAGC 7572  
 QY 628 TrpThrSerIleAsp 632  
 Db 7573 TGGACTGCGATGAAC 7587  
 RESULT 9  
 AAF80285  
 ID AAF80285 standard; DNA; 8654 BP.  
 AC AAF80285;  
 XX 29-JUN-2001 (first entry)  
 DT 29-JUN-2001 (first entry)  
 XX Nucleotide sequence of plasmid pMR1192.  
 DE Vector; transgenesis; ttfA locus; RK2 ori; oriV; p285 protein;  
 KW P382 protein; antibiotic resistance gene; nptII; transgenic plant; ss.  
 XX Synthetic.  
 OS  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT 1..654  
 FT /tag= a  
 FT /note= "ori RK2"  
 FT 655..1263  
 FT /tag= b  
 FT /note= "ori ColE1"  
 FT 1264..2603  
 FT /tag= c  
 FT /note= "NPT III gene coding for neomycin  
 FT phosphotransferase and kanamycin resistance"  
 FT 2604..4098  
 FT /tag= d  
 FT /note= "ttrA locus from RK2 coding for two proteins P285

and p382 enabling an increase in the replication rate"

FT misc\_feature /tag= e  
FT 4106..4271  
FT /note= "T-DNA left border"  
FT terminator 4272..4559  
FT /tag= f  
FT /note= "nopaline synthetase terminator"  
FT CDS 4560..5556  
FT /tag= g  
FT /note= "NPT III gene coding for neomycin  
FT phosphotransferase and kanamycin resistance"  
FT promoter 5557..5771  
FT /tag= h  
FT /note= "nopaline synthetase promoter"  
FT CDS 5818..7717  
FT /tag= i  
FT /note= "GUS gene coding for beta glucuronidase"  
FT polyA\_signal 7718..8447  
FT /tag= j  
FT /note= "polyA from 35S ribosome"  
FT mic\_feature 8474..8647  
FT /tag= k  
FT /note= "T-DNA right border"

PN FR2798139-A1.  
XX 09-MAR-2001.  
XX 03-SEP-1999; 99FR-00011112.  
XX 03-SEP-1999; 99FR-00011112.  
XX (MERIT-) MERISTEM THERAPEUTICS SA.  
XX Gruber V, Comeau D;  
XX WPI; 2001-259847/27.  
XX  
XX New vector free from non-essential elements, useful for transforming  
XX cells for protein production and for preparing transgenic plants.

PS Claim 20; Page 119-122; 180pp; French.

CC The specification describes a synthetic vector containing only those  
CC elements essential for its functionality and transgenesis of a cell  
CC (especially a plant cell). The vector consists of at most one origin of  
CC replication (ori), at most one sequence encoding a selection agent and a  
CC trfA locus encoding a protein that increases the level of plasmid  
CC replication. The vector particularly contains an RK2 ori, especially oriV  
CC from pRK2 of Escherichia coli with a broad host range, an antibiotic  
CC resistance gene (especially nptII conferring resistance to kanamycin in  
CC bacteria) and a trfA locus from pRK2 encoding the proteins P285 and P382.  
CC The vectors are used to prepare transgenic plants and transformed host  
CC cells for production of a heterologous proteins, e.g. insulin,  
CC interferon, lipase, blood proteins and anti-inflammatory agents. The  
CC present sequence represents a plasmid of the invention

XX  
SQ Sequence 8654 BP; 2242 A; 2134 C; 2352 G; 1926 T; 0 U; 0 Other;

Alignment Scores:  
Pred. NO.:  
Score: 3.68e-138 Length: 8654  
Percent Similarity: 1714.50 Matches: 337  
Best Local Similarity: 69.12% Conservative: 95  
Query Match: 53.92% Mismatches: 177  
DB: 51.12% Indels: 17  
Gaps: 9

US-10-757-093-4 (1-634) x AAF80285 (1-8654)

Gy 15 ProserLNGUjYTHPrCoALaLaRghIstPheProArgaNgUmEtThrcInHieglu 34  
Db |||||CGGAAGCTGGCGGCCGTTCACACGCCTGGATCCT--AAATTAGTGCACTTAAGGGGT 5799  
5742 CGGAAGCTGGCGGCCGTTCACACGCCTGGATCCT--AAATTAGTGCACTTAAGGGGT 5799

[illegible]

```

Db      6853 GTGATGATGAAACTGCTGCTGCTTAACTCTTAGCATTTGGTTGAAAGCG 6912
Qy      409 SerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAla 427
Db      6913 GCGCAACAGCGGAAAGCTGATCAGCAAGAGGAGCACTCAACGCGGAACCTCAGCAACGCG 6972
Qy      428 HisLysGlnAlaIleAsnGluLeuIleAlaArgAspLysAsnHisAlaSerValValMet 447
Db      6973 CACTTACGGCGGATTAAAGACTATAGCCGCTACAAAACCAACCCAGAGCGTGGTATG 7032
Qy      448 TrpSerIleAlaAsnGluProAlaSerHisGlnAspGlyAlaArgGluTrpPheGluPro 467
Db      7033 TGGAGTATGTCGAACGAACCGGATACCCGCCGAGGTGACGAGATATTTGCGCCA 7092
Qy      468 LeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGly 487
Db      7093 CTGGCGGAGCAACGCGGTAACTGACCCGACCGGCTCGATCACCTGCTCAATGTATG 7152
Qy      488 ThrAlaThrTrpGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArg 507
Db      7153 TTCTGCGACGCTCACACCGATACCATCAGCATCTTTATATGTGCTGCTGAAACCGT 7212
Qy      508 TyrPheGlyTrpTrpSerGlnThrGlyAspLeuGlnGluAlaGluAlaIleGluLys 527
Db      7213 TAATTACGATGTATGTATGTCAAAGCGCGGATTTGGAAACGCGAGAGAGTACTGAAAAA 7272
Qy      528 GlnLeuHisGlyTrpGlnGlnLysPheHisArgProIleValMetThrGluTrpGlyAla 547
Db      7273 GAACCTTGGCGCTGGCAGAGAACTGATCAGCCGATTCATCACCGAATACGGGGTG 7332
Qy      548 AspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGlnGluPheGlnVal 567
Db      7333 GATACGTTACCCCGGCTGCACTCAATGATACCGACACATGTGAGTGAAGATCACTGT 7392
Qy      568 GlnMetLeuAspMetTrpHisArgValPheAspArgIleGluSerMetAlaGlyGlnHis 587
Db      7393 GCATGGCTGATGTATGATCAACCGGCTTGTGATCGCTCAGCCGCTGCGGTGAACAG 7452
Qy      588 ValTrpAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnLys 607
Db      7453 GTATGCAATTTCCCGCATTTTGGACCTCGCAAGGCAATTTGGCGGTTGCGGTAACAG 7512
Qy      608 LysGlyValPheThrArgAspArgLysProLysAlaIleAlaHisSerLeuArgAlaArg 627
Db      7513 AAAGGATCTTCACTGCGCAACCGAAACGAAATCGCGGCTTTCTGCTGCAAAAAACGC 7572
Qy      628 TrpThrSerIleAsp 632
Db      7573 TGGACTGGCATGAAC 7587

RESULT 10
AAFe0296
ID      AAF80296 standard; DNA; 8987 BP.
XX      AAF80296;
XX      29-JUN-2001 (first entry)
XX      Nucleotide sequence of plasmid pMR11212.
XX      DE
XX      Vector; transgenesis; trfA locus; RK2 ori; oriV; P285 protein;
XX      KW P382 protein; antibiotic resistance gene; nptIII; transgenic plant; ss.
XX      OS
XX      Synthetic.
XX      FH
XX      Key      Location/Qualifiers
XX      rep_origin      1..654
XX      /tag= a
XX      /note= "ori RK2"
XX      rep_origin      655..1263
XX      /tag= b
XX      /note= "ori ColE1"

```

```

FT      CDS      1264..2603
FT      /tag= c
FT      /note= "NPT III gene coding for neomycin
FT      phosphotransferase and kanamycin resistance"
FT      CDS      2604..4098
FT      /tag= d
FT      /note= "trfA locus from RK2 coding for two proteins P285
FT      and P382 enabling an increase in the replication rate"
FT      terminator 4272..4459
FT      /tag= e
FT      /note= "nopaline synthetase terminator"
FT      CDS      4575..5150
FT      /tag= f
FT      /note= "Bar gene coding for phosphinotricin
FT      acetyltransferase and glutosinate resistance"
FT      promoter 5151..5368
FT      /tag= g
FT      /note= "nopaline synthetase promoter"
FT      promoter 5369..6111
FT      /tag= h
FT      /note= "enhanced promoter from 35S ribosome"
FT      CDS      6159..8050
FT      /tag= i
FT      /note= "gus gene coding for beta glucuronidase"
FT      polyA_signal 8051..8780
FT      /tag= j
FT      /note= "polyA from 35S ribosome"
FT      misc_feature 8807..8980
FT      /tag= k
FT      /note= "T-DNA right border"
FT      FN298139-A1.
FT      09-MAR-2001.
FT      03-SEP-1999; 99FR-00011112.
FT      03-SEP-1999; 99FR-00011112.
FT      (MER1-) MERISTEM THERAPEUTICS SA.
FT      Gruber V, Comeau D;
FT      MPI; 2001-259847/27.
FT      New vector free from non-essential elements, useful for transforming
FT      cells for protein production and for preparing transgenic plants.
FT      Claim 20; Page 168-171; 180pp; French.
XX      The specification describes a synthetic vector containing only those
XX      elements essential for its functionality and transgenesis of a cell
XX      (especially a plant cell). The vector consists of at most one origin of
XX      replication (ori), at most one sequence encoding a selection agent and a
XX      trfA locus encoding a protein that increases the level of plasmid
XX      replication. The vector particularly contains an RK2 ori, especially oriV
XX      from pRK2 of Escherichia coli with a broad host range, an antibiotic
XX      resistance gene (especially nptIII conferring resistance to kanamycin in
XX      bacteria) and a trfA locus from pRK2 encoding the proteins P285 and P382.
XX      The vectors are used to prepare transgenic plants and transformed host
XX      cells for production of a heterologous proteins, e.g. insulin,
XX      interferon, lipase, blood proteins and anti-inflammatory agents. The
XX      present sequence represents a plasmid of the invention
XX      SO      Sequence 8987 BP; 2342 A; 2205 C; 2438 G; 2002 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      3,87e-138
Score:          1714.50
Percent Similarity: 69.12%
Best Local Similarity: 53.92%
Query Match:    51.12%
DB:             4
Gaps:           9
Length:         8987
Matches:        337
Conservative:   95
Mismatch:       177
Indels:         17
Gaps:           9

```

```

US-10-757-093-4 (1-634) x AAF80296 (1-8987)
QY 15 ProSerLeuGlyThrProAlaAlaArgHisPheProArgAsnGluMetThrGlnHisGlu 34
DB 6075 CCCAAGCTTGGCCGGCGGTAAACACGGCTGGATCTT--AATTAAAGTCAGCTTAGGGGT 6132
QY 35 GlnProLeuIleuValArgProGlnArgThrSerSerArgProGluLeuValAsnLeuAsp 54
DB 6133 GGTCAAGTCCCTTATGTAAGCTCTGTAGAAACCCCAACCGGTGAATCAAAAAAATCTGAC 6192
QY 55 GlyLeuTrpIlePheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnPro 71
DB 6193 GGGCTGGGGCATTCAGCTCGATCGGAAACCTGTGAATTGATCAG-----CGTTGG 6246
QY 72 TrpThrAlaProLeuProIleGlyLeuGluCybProValProAlaSerTrpAsnAspIle 91
DB 6247 TGGGAAAGCCGGTTACAAAGAAAGCCGGCAATTGCTGCGAGCGAGTTTAAACGATCAG 6306
QY 92 PheIleSerArgGluIleHisAspHisValGlyTrpValTrpTrpGlnArgGluValIle 111
DB 6307 TTGGCCGATGCAAGATATTCTGTATATGCGGCAACGCTGTGTATCAGCCGAAAGTCTTT 6366
QY 112 ValProIleGlyTrpSerGlnArgTrpLeuValArgAlaGluSerAlaThrHisHis 131
DB 6367 ATACGAAAGGTGGGAGCGCAGCGCATGCTGCTGCTTCATGCGGTCACTCATTAC 6426
QY 132 GlyArgIleTrpValAsnAsnArgLeuValAlaGlnHisValGlyGlyTrpProPhe 151
DB 6427 GGCAGAAAGTGGCGCATTAATCAAGAAAGTATGAGCATGAGCGGCTTACGCCCATTT 6486
QY 152 GlnAlaAspValThrGluLeuValAlaProGlyGluIlePheArgLeuThrIleGlyVal 171
DB 6487 GAAGCCCATGTCACCGCTATGTTATGTCGCGGAAAGTGTACCTATCACCGTTGTG 6546
QY 172 AsnAsnGluLeuThrHisGluThrIleProProGlyIleTrpHisLeuAsnAlaThr 191
DB 6547 AACACAGAACTGAACCTGCAGACTATCCCGCGGAATGGATTAAC--GACGAAAC 6603
QY 192 GlyLeuArgIleGlnThrTrpGlnHisAspPheTrpAsnTrpAlaGlyLeuAlaArgSer 211
DB 6604 GGCAGAAAGAGCGCTTCACTTCATGATTTCTTAACTATGCGGAATCCATCCGACG 6663
QY 212 IleTrpLeuTrpSerValProGlnGlnHisIleGlnAspIleThrValIleThrAspVal 231
DB 6664 GTATGCTGTACACACGCGCAACACCTGGGTGACATATCACCGGTGACCATGTC 6723
QY 232 -----AspGlyAspAsnGlyLeuIleAsnTrpGluValGluValAlaAsnGlnThrThr 249
DB 6724 GCCCAGACCTGTAAACACCGGCTGTGACTGCGAGGTG---GTGGCCAT----- 6771
QY 250 GlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAlaIleValAsnGly 269
DB 6772 GGTGATGTACAGCTGTGAACGCGGTGATGCGGATCAACAGGTGGTTGCACTGCACAGGC 6831
QY 270 AlaGlnGlyThrValThrIleProSerValIleValLeuTrpGlnProGlyAlaIleVal 289
DB 6832 ACTAGCGGAGCTTTGCAAGTGGTGAATCCGCACTGCGCAACCGGCTGAAGGTTATCTC 6891
QY 290 TrpGlnLeuGlnValAsnIleValIleGlySerSerGlyAspValValAspThrTrpAsnLeu 309
DB 6892 TATGAACCTGGCGTC-----ACAGCCAAAGCCAGACAGATGATATCTACCCGCTT 6945
QY 310 AlaThrGlyValArgThrValIleValAlaGlySerGlnPheLeuIleAsnGlyLeuPro 329
DB 6946 CGCGTCGGCATCCGCTAGTGGCAGTGAAGGCGCAACAGTTCCTGATTAACCAAAACCG 7005
QY 330 PheTrpPheThrGlyPheGlyLeuHisGlyAspThrAlaValArgGlyLeuGlyHisAsp 349
DB 7006 TTCTACTTACTGGCTTGGTGGTCATGAGATGCGCACTTGGCGCAAGGATTGAT 7065
QY 350 ProAlaTrpMetValHisAspPheGlnLeuMetLeuTrpIleGlyAlaAsnSerPheArg 369

```

```

DB 7066 AACGTGCTAGTGGTGCACGACGATTAATGACTGTGAGGCGCAACTCTACCGT 7125
QY 370 ThrSerHisTrpProTrpAlaGluGluValMetAspPheAlaAspArgAsnGlyIleVal 389
DB 7126 ACTCTGCATTAACCTTACGCTGAAGATGCTGACTGGGCAATGAACATGGCATCTGT 7185
QY 390 ValIleAspGluThrProAlaValIleGlyLeuAsnIleAlaLeu--MetGlyValSerGlu 408
DB 7186 GTGATGATGAACCTGCTGCTGCTGCTTAACTCTCTTAAAGCATTTGTTTGAAGCG 7245
QY 409 SerGlyAlaPro--GlnThrPheThrProAspAlaIleAsnAspIleThrGlnGluAla 427
DB 7246 GGCACACAGCCGAAAGAACTGTACAGCGAAAGCGAGTCAACCGGAAACTCAGCAAGCG 7305
QY 428 HisLeuGlnAlaIleArgGluLeuIleAlaArgAspIleAsnHisIleAsnSerValIleMet 447
DB 7306 CACTTACAGCGGATTAAGCTGTGAATGCCGCTGACCAAAAAACCAACCAACCGGTGATG 7365
QY 448 TrpSerIleAlaAsnGluProAlaSerHisGlyLeuAspGlyValAlaArgGlyTrpPheGluPro 467
DB 7366 TGGAGTATTGCCAACGACCGGATACCCGTCGCCCAAGTGCACGGAAATTTTCGGCCA 7425
QY 468 LeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGly 487
DB 7426 CTGCGGAAACCAACGGGTAACTGCAACCGCAGCGCTCCGATCACTCCGCTCAATGTATG 7485
QY 488 ThrAlaThrTrpGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArg 507
DB 7486 TTCTGCACGCTCACACCGATACCATCAGCATCTTGTGATGCTGTGCTGAAACCGT 7545
QY 508 TrpPheGlyTrpTrpSerGlnThrGlyAspLeuGluGluAlaIleAlaLeuGluIleVal 527
DB 7546 TATTACGATGATGATGATGCCAACCGCGATTTGGAAACGCAAGAAAGTACTGGAATA 7605
QY 528 GlyLeuHisGlyTrpGlnGluIlePheHisAspProIleValMetThrGlyTrpGlyAla 547
DB 7606 GAACCTTCTGCGCTGGCAGGAGAACTGCATCACCGCATTTCACTACCGGAATCGGCTG 7665
QY 548 AspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGluGluPheGlnVal 567
DB 7666 GATACGTTACCGCGGCTGCATCAATGTACACCGACATGCGAGTGAAGATGACATGT 7725
QY 568 GlnMetLeuAspMetTrpHisArgValPheAspArgIleGluSerMetAlaGlyGlnHis 587
DB 7726 GCATGCTGATATGATGATCCCGCTTTGATCGGCTGACGCCGCTGCGGTGAACAG 7785
QY 588 ValTrpAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnIleVal 607
DB 7786 GTATGGAATTTGCCGATTTTGGACCTGCAAGGCATATTTGCGGTTGGCGGTAAACAAG 7845
QY 608 IysGlyValPheThrArgAspArgIleProIleValAlaIleHisSerLeuAlaGlyAlaArg 627
DB 7846 AAAGGATCTTCACTGCGCAACCGAAACGAAAGTCCGCGCTTTTCTGTCAAAACGCG 7905
QY 628 TrpThrSerIleAsp 632
DB 7906 TGGATCGCATGAAC 7920

```

RESULT 11  
 AAF80292  
 ID AAF80292 standard; DNA; 9390 BP.  
 AC AAF80292;  
 DT 29-JUN-2001 (first entry)  
 XX  
 DE Nucleotide sequence of plasmid pMRT1204.  
 KW Vector; transgenesis; teta locus; RK2 ori; p285 protein;  
 KW P382 protein; antibiotic resistance gene; mpcIII; transgenic plant; ss.  
 XX  
 OS Synthetic.

	Location/Qualifiers
FT Key	1..654
FT rep_origin	/*tag= a
FT	/note= "Ori RK2"
FT	655..1263
FT rep_origin	/*tag= b
FT	/note= "ori ColEI"
FT	1264..2603
FT CDS	/*tag= c
FT	/note= "NPT III gene coding for neomycin phosphotransferase and kanamycin resistance"
FT	2604..4098
FT CDS	/*tag= d
FT	/note= "TrfA locus from RK2 coding for two proteins P285 and P382 enabling an increase in the replication rate"
FT	4106..4271
FT misc_feature	/*tag= e
FT	/note= "T-DNA left border"
FT	4272..4559
FT terminator	/*tag= f
FT	/note= "nopaline synthetase terminator"
FT	4560..5556
FT CDS	/*tag= g
FT	/note= "NPT III gene coding for neomycin phosphotransferase and kanamycin resistance"
FT	5560..5771
FT promoter	/*tag= h
FT	/note= "nopaline synthetase promoter"
FT	5772..6514
FT promoter	/*tag= i
FT	/note= "enhanced promoter from 35S ribosome"
FT	6554..8453
FT CDS	/*tag= j
FT	/note= "GUS gene coding for beta glucuronidase"
FT	8454..9183
FT polyA_signal	/*tag= k
FT	/note= "polyA from 35S ribosome"
FT	9210..9383
FT misc_feature	/*tag= l
FT	/note= "T-DNA right border"
FR2798139-Al.	
PD 09-MAR-2001.	
XX 03-SEP-1999;	99FR-00011112.
XX 03-SEP-1999;	99FR-00011112.
XX 03-SEP-1999;	99FR-00011112.
XX (MERI-) MERISTEM THERAPEUTICS SA.	
XX PA	
XX PI	
XX Gruber V, Comeau D;	
XX WPI; 2001-259847/27.	
XX	
XX New vector free from non-essential elements, useful for transforming	
XX cells for protein production and for preparing transgenic plants.	
XX	
XX Claim 20; Page 149-152; 180bp; French.	
XX	
XX The specification describes a synthetic vector containing only those	
XX elements essential for its functionality and transgenesis of a cell	
XX (especially a plant cell). The vector consists of at most one origin of	
XX replication (ori), at most one sequence encoding a selection agent and a	
XX trfA locus encoding a protein that increases the level of plasmid	
XX replication. The vector particularly contains an RK2 ori, especially oriV	
XX from pRK2 of Escherichia coli with a broad host range, an antibiotic	
XX resistance gene (especially nptII conferring resistance to kanamycin in	
XX bacterial) and a trfA locus from pRK2 encoding the proteins P285 and P382.	
XX The vectors are used to prepare transgenic plants and transformed host	
XX cells for production of a heterologous proteins, e.g. insulin,	
XX interferon, lipase, blood proteins and anti-inflammatory agents. The	
XX present sequence represents a plasmid of the invention	

XX	Sequence	9390 BP; 2478 A; 2322 C; 2510 G; 2080 T; 0 U; 0 Other;	
	Alignment Scores:		
	Pred. No.:	4, 1e-138	9390
	Score:	1714.50	337
	Percent Similarity:	69.12%	Conservative: 95
	Best Local Similarity:	53.928	Mismatches: 177
	Query Match:	51.12%	Indels: 17
	DB:	4	Gaps: 9
	US-10-757-093-4 (1-634) x AAF80292 (1-9390)		
OY	15	ProSerLeuGIYThrProAlaAlaArgHisPheProAlaGlnMeThcGlnIsclu	34
Db	6478	CCCAAGCTGGCGCGCGCGGCTTAACAGCGCGTGCATCTT--AAATTAAGTCGACTTAAGGGGT	6535
OY	35	GlnProLeuIleLysValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAsp	54
Db	6536	GGTCAGTCCCTTAATGCTTAAGTCTCTGAGAAACCCCAACCCCGTAATCAAAAACCTGCAC	6595S
OY	55	GIYLeuTrpLysPheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnPro	71
Db	6596	GGCTGTGGCGCATTCAGTGTGGATGCGCAAAAACCTGTGAATGTATTCAG-----CGTTGG	6649
OY	72	TrpThrAlaProLeuProLysGlyLeuGluCysProValProAlaSerTyrAsnAspIle	91
Db	6650	TGGGAAAGCGGCTTACAGAAAGCCGGGCATTCGCTGTCCAGCGCATTTTAACGATCAG	6709
OY	92	PheIleSerArgGluIleHisAspHisValGlyTyrValTyrGlnAspGluValIle	111
Db	6710	TTGGCCGATGCAGATATTCGTATTAATTCGGGCAACGCTGGTATTCAGCGCGGAAGTCTTT	6769
OY	112	ValProLysGlyTyrPheSerGlnLysArgTyrLeuValArgAlaGlnSerAlaThrHisHis	131
Db	6770	ATACCGAAAGGTGGGCGAGCCAGCGATATGCTGTGCGTTTCATTCGGGCATCATTAC	6829
OY	132	GlyArgIleTyrValAsnAspArgLeuValAlaGlnHisValGlyGlyTyrThrProPhe	151
Db	6830	GGCAAAAGTGGGTCAATATTCAGAGATGTGTGAGCATCAGGGCGGCTTATACGCCATTTT	6889
OY	152	GluAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyVal	171
Db	6890	GAAAGCGATGCACGCCGTATGTTATTCGGGGAAGAGTACGTATCCATCCGTTTGCTG	6949
OY	172	AspAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThr	191
Db	6950	AACAACGAACTGAACTGGACAGACATATCCCGCGGAATGTGTATTC---GACGAAAC	7006
OY	192	GlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSer	211
Db	7007	GGCAAGAAAAAGCAGTCTTACTTCATGATTTCTTTAATCAATACCCGGAATCCATCCGACG	7066
OY	212	IleThrLeuTyrSerValProGlnGlnHisIleGlnAspIleThrValValThrAspVal	231
Db	7067	GTAATGCTCTTACACCAACCGGAAACACCTGGGTGAGATATCACCGTGTGACGCATGTC	7126
OY	232	-----AspGlyAspAsnGlyLeuIleAsnTyrGluValGluValAlaAsnGlnThrThr	249
Db	7127	GCGCAACAGCTGAACCAACGCGGTCTGTGACTGGCAAGT---GTGGCCAAAT-----	7174
OY	250	GlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAlaLysAlaSerGly	269
Db	7175	GGGTATGTCAGCGTTGAACCTGCTGATGCGGATCAACAGTGTGTCACACTGGACAAAGCG	7234
OY	270	AlaGlnIleThrValThrIleProSerValLysLeuThrGlnProGlyValAlaTyrLeu	289
Db	7235	ACTAAGCGGACCTTGCAGATGGTGAATCCGCACCTGTGGCAACCGGCTAAGGTTATATTC	7294
OY	290	TyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValValAspThrTyrAsnLeu	309
Db	7295	TATGAACTGTGCGTTC-----ACAGCGAAAGCGCAGACGAGTGGATATCTTACCGGCTT	7348



```

QY 310 AlaThrGlyValArgThrValIleValIleGlySerGlnPheLeuIleAsnGlyLysPro 329
DB 7349 CCGCGTGGCAATCCGGTCAAGTGCATGAAAGGGCAAGTTCTTGATTAACCAACCG 7408
QY 330 PheTYrPheThrGlyPheGlyLysHISGlyAspThrAlaValArgIleLysGlyHISAsp 349
DB 7409 TTCTACTTACTGCTTGGTGTGCTCATGAAGATCGGACCTGCGTGGCAAGATTCGAT 7468
QY 350 ProAlaTYrMetValHISAspPheGlnLeuMetLysTrpIleGlyAlaAsnSerPheArg 369
DB 7469 AACGTGCTGATGTTGTCACGACCAAGCATTAATGATGATGTTGGGGCAACTCTTCCTG 7528
QY 370 ThrSerHISTYrProTYrAlaGluGluValMetAspPheAlaAspArgAsnGlyIleVal 389
DB 7529 ACCTGCAATTAACCTTAACCTGTAAGAGATGCTGATGGGCAATGAACTGGCATTCGTG 7588
QY 390 ValIleAspGluThrProAlaValIleGlyLeuAsnIleAlaLeu--MetGlyValSerGlu 408
DB 7589 GTGATTGATGAACCTGCTGCTGCGCTTACCTCTTTAGCATTTGGTTTCGAACCG 7648
QY 409 SerGlyAlaPro--GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAla 427
DB 7649 GGCACAAAGCCGAAACCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7708
QY 428 HisLysGluAlaIleArgGluLeuIleAlaArgAspLysAsnHISAlaSerValIleMet 447
DB 7709 CACTTACAGGCGGATTAAGAGCTGATAGCCGCTGACAAACCAACCAACCGTGGTATG 7768
QY 448 TrpSerIleAlaAsnGluProAlaSerHISGlyAspGlyValArgIleTYrPheGluPro 467
DB 7769 TGGAGATTTGCCAAGCAACCGATACCCGCGCAAGTGTACAGGGAAATTTTCGGCCA 7828
QY 468 LeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGly 487
DB 7829 CTGGCCGAAAGCAACCGGTAACCTGACCCGACCGCTCGATCACTCGCGCAATGTAATG 7888
QY 488 ThrAlaThrTYrGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArg 507
DB 7889 TTCTGGACGCTACACCGATACATCAGGATCTTTGATGTCGTGCTGACCGT 7948
QY 508 TYrPheGlyTYrTYrSerGlnThrGlyAspLeuGluGluAlaIleAlaLeuGlyLys 527
DB 7949 TATTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8008
QY 528 GluLeuHISGlyTYrGlnGlyLysPheHISArgProIleValMetThrGlyTYrGlyAla 547
DB 8009 GAACCTTCTGGCTGCGCAGAGAACTGCATCAGCCGATTTATCATCAACCGAATACGG 8068
QY 548 AspThrLeuAlaGlyLeuHISerIleLeuGlyLeuProTrpSerGluGluPheGlnVal 567
DB 8069 GATACGTTAGCCGGGCTGCACTCATATGATACCGACATGTCAGTGAAGATGATGATG 8128
QY 568 GlnMetLeuAspMetTYrHISArgValPheAspArgIleGluSerMetAlaGlyLysHIS 587
DB 8129 GCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8188
QY 588 ValTrpAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnLys 607
DB 8189 GATATGAAATTTCCCGATTTTGGACCTTCGCAAGGCAATATTTGGCGGTGATCAAG 8248
QY 608 LysGlyValPheThrArgAspArgLysProLysAlaIleAlaHISerLeuArgAlaArg 627
DB 8249 AAGGGAATCTTCACTGCGCAACCGAAGATGCGGCGGCTTTCTGCTGCAAAAACGC 8308
QY 628 TrpThrSerIleAsp 632
DB 8309 TGCACCTGCAATGAAC 8323

```

```

XX 29-JUN-2001 (first entry)
DT Nucleotide sequence of plasmid pMR1206.
XX Vector; transgeneis; trfA locus; RK2 ori; oriV; P285 protein;
XX P282 protein; antibiotic resistance gene; nptIII; transgenic plant; ss.
XX Synthetic.
OS
FH Key
FH rep_origin
FT Location/Qualifiers
FT 1..654
FT /tag= a
FT /note= "ori RK2"
FT 655..1263
FT /tag= b
FT /note= "ori ColE1"
FT 1264..2603
FT /tag= c
FT /note= "npt III gene coding for neomycin
FT phosphotransferase and kanamycin resistance"
FT 2604..4098
FT /tag= d
FT /note= "trfA locus from RK2 coding for two proteins P285
FT and P282 enabling an increase in the replication rate"
FT 4106..4271
FT /tag= e
FT /note= "T-DNA left border"
FT 4272..4559
FT /tag= f
FT /note= "nopaline synthetase terminator"
FT 4560..5559
FT /tag= g
FT /note= "npt III gene coding for neomycin
FT phosphotransferase and kanamycin resistance"
FT 5560..5771
FT /tag= h
FT /note= "nopaline synthetase promoter"
FT 5772..6514
FT /tag= i
FT /note= "enhanced promoter from 35S ribosome"
FT 6554..8453
FT /tag= j
FT /note= "GUS gene coding for beta glucuronidase"
FT 8454..9183
FT /tag= k
FT /note= "polyA from 35S ribosome"
FT 9210..9383
FT /tag= l
FT /note= "T-DNA right border"
PN FR2798139-A1.
XX 09-MAR-2001.
PD
XX 03-SEP-1999; 99FR-00011112.
XX 03-SEP-1999; 99FR-00011112.
XX (MERI-) MERISTEM THERAPEUTICS SA.
XX Gruber V, Comeau D;
XX WPI, 2001-259847/27.
XX New vector free from non-essential elements, useful for transforming
XX cells for protein production and for preparing transgenic plants.
XX Claim 20; Page 158-161; 180pp; French.
XX The specification describes a synthetic vector containing only those
XX elements essential for its functionality and transgeneis of a cell
XX (especially a plant cell). The vector consists of at most one origin of

```

CC replication (ori), at most one sequence encoding a selection agent and a  
 CC triA locus encoding a protein that increases the level of plasmid  
 CC replication. The vector particularly contains an RK2 ori, especially oriV  
 CC from *PRK2* of *Escherichia coli* with a broad host range, an antibiotic  
 CC resistance gene (especially *mpcIII* conferring resistance to kanamycin in  
 CC bacteria) and a triA locus from *PRK2* encoding the proteins p285 and p382.  
 CC The vectors are used to prepare transgenic plants and transformed host  
 CC cells for production of a heterologous proteins, e.g. insulin,  
 CC interferon, lipase, blood proteins and anti-inflammatory agents. The  
 CC present sequence represents a plasmid of the invention

XX Sequence 9390 BP; 2479 A; 2322 C; 2510 G; 2079 T; 0 U; 0 Other;

# Alignment Scores:

Pred. No.:	Length:	9390
Score:	1714.50	Matches: 337
Percent Similarity:	69.12%	Conservative: 95
Best Local Similarity:	53.92%	Mismatches: 177
Query Match:	51.12%	Indels: 17
DB:	4	Gaps: 9

US-10-757-093-4 (1-634) x AAF80294 (1-9390)

QY 15 ProSerLeuGlyThrProAlaAlaArgHisPheProArgAsnGluMetThrGlnHisGlu 34  
 DB 6478 CCCAAGCTTGGCCGCGCTTAACACGGGTGATCCTT--AATTAAAGTCGACTTAGGGGT 6535  
 QY 35 GlnProLeuLeuValArgProGlnArgTherSerArgGluLeuValAsnLeuAsp 54  
 DB 6536 GGTGACGCTTATGTTACGTCCTGTAAGAAACCCCAACCGCTGAATCAAAAAAAGCTGAC 6595  
 QY 55 GlyLeuTrpLysPheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnPro 71  
 DB 6596 GGGCTGTGGCATTCACTGCTGATCGCGAAACTGTGTAATGATCAG-----CGTTGG 6649  
 QY 72 TrpThrAlaProLeuProLysGlyLeuGluCysProValProAlaSerGlyTrpAsnAsp 91  
 DB 6650 TGGGAAAGCCGTTACAGAAAGCCGGGCAATTGCTGTGCCAGCGATTTAACGATCAG 6709  
 QY 92 PheIleSerArgGluLeuHisAspHisValGlyTrpValArgGlnArgGluVal 111  
 DB 6710 TTGCGCGATGACATATTCGTAATTAATGCGGCAAGCTGATACGCCGAAGCTTTT 6769  
 QY 112 ValProLysGlyTrpSerGlnGluArgGlyLeuValArgAlaGluSerAlaThrHis 131  
 DB 6770 ATACCAAGAGTTGGGAGGCGGCGATTCGTCGCTTCGATCGGTCCTCATTTAC 6829  
 QY 132 GlyArgIleTyrValAsnAsnArgLeuValAlaGluHisValGlyGlyTrpThrProPhe 151  
 DB 6830 GCGAAGTGTGGCTCAATATACGAAAGTATGAGACATCAGGCGGCTAATACGCCATT 6889  
 QY 152 GluAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyVal 171  
 DB 6890 GAAGCGGATGTCAGCGGATGTTATTTGCGGGAAAGTACGATCAACCGTTGTGTG 6949  
 QY 172 AsnAsnGluLeuThrHisGluTrpIleProProGlyLysIleThrThrGlyAsnAlaThr 191  
 DB 6950 AACCAAGAACTGAACCTGCGACTATCCGCGGGAGATGAGTATTAAC---AGCAAAAAC 7006  
 QY 192 GlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSer 211  
 DB 7007 GCGAAGAAAAGAGCTTACTTCCATGATTTCTTAACTATCCCGGAATCCATGCGAAC 7066  
 QY 212 IleTrpLeuTyrSerValProGlnGlnHisIleGlnAspIleThrValValThrAspVal 231  
 DB 7067 GTAATGCTCTTACACCGCCGAACACCTGGGTGACAGATATCACCGTGTGACGATGTC 7126  
 QY 232 -----AspGlyAspAsnGlyLeuIleAsnTyrGluValAlaGluValAlaAsnGlnThr 249  
 DB 7127 GCGCAAGACTGTAAACACCGCTGTGTGACTGCGCGGTG--GTGGCCAAAT----- 7174  
 QY 250 GlyIleIleGlnIleSerValIleAspGluAspGlyAlaIleValAlaIleValAlaSerGly 269

DB 7175 GGTGATGTCAGCTTGAACCTGCTGATGCGGATCAACAGGTGTTCAACTGACAGGC 7234  
 QY 270 AlAGlnGlyThrValThrIleProSerValLysLeuTrpGlnProGlyValAlaTyrLeu 289  
 DB 7235 ACTAAGGGGACTTTGGAGAGTGAATTCGCACACTCTGCGCAACCGGGGTGAAGTTATCTC 7294  
 QY 290 TyrGlnLeuGlnValAsnIleValGlySerSerGlyAspValAlaAspThrThrAsnLeu 309  
 DB 7295 TATGAACGTGGCTC-----ACAGCCAAAGCCAGACAGCTGTATATCTAACCGCTT 7348  
 QY 310 AlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysPro 329  
 DB 7349 CGCGTCGCGATCCGGTCAGTGGAGTGAAGGCGAACAATCTCTATTAACCAAAACCG 7408  
 QY 330 PheTyrPheThrGlyPheGlyLysHisGlyAspThrAlaValArgGlyLysGlyHisAsp 349  
 DB 7409 TTCTACTTTCATCGGCTTGTGCTGATGAAAGTGGAGACTTCTGCGCAAAAGATTGAT 7468  
 QY 350 ProAlaTyrMetValHisAspPheGlnLeuMetLysTrpIleGlyValAsnSerPheArg 369  
 DB 7469 AACGTGCTGATGGTGCACACACCGCATTAATGACTGGATTGGGCGCAACTCTTACCGT 7528  
 QY 370 ThrSerHisTyrProTyrAlaGluGluValMetAspPheAlaAspArgAsnGlyIleVal 389  
 DB 7529 ACCTGCGATTACCTTACGCTGAAGAGATGCTGACGTGGCAGATGAACATGGCATCGTG 7588  
 QY 390 ValIleAspGlnThrProAlaValGlyLeuAsnIleAlaLeu--MetGlyValSerGlu 408  
 DB 7589 GTGATTGATGAAGACCTGCTGCTGCTTAACTCTTTAAGCATGGTTTCGAAACCG 7648  
 QY 409 SerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluVal 427  
 DB 7649 GGCACAAAGCCGAAAGAACTGTACAGCAAGAGAGCAACAGGGGAAACTCAGCAACCG 7708  
 QY 428 HisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMet 447  
 DB 7709 CACTTAACGGCGATTAAGAGCTGATAGCGCTGCAAAAACCAACCAAGCGTGGTATG 7768  
 QY 448 TrpSerIleAlaAsnGluProAlaSerHisGluAspGlyValAspGlyLysTrpPheGluPro 467  
 DB 7769 TGGAGTATTTGCCAAGAACCGGATACCGCTCGCAAGGTGACGGGAATATTTGCGGCA 7828  
 QY 468 LeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGly 487  
 DB 7829 CTGGGGAAGCAACGCTGAACCTGACCGAGCGGTGCAACCTGCTCAATGATATG 7888  
 QY 488 ThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArg 507  
 DB 7889 TTGCGCAGCGTCAACCGCATATCCATCGCATCTTTGATGTGCTGCTGCAACCGT 7948  
 QY 508 TyrPheGlyTyrTyrSerGlnThrGlyAspLeuGluGluAlaGluAlaIleGluLys 527  
 DB 7949 TATTACGATGTATATGTCACAAAGCGGATTTGGAACCGCAAGAAAGTACTGGAATAA 8008  
 QY 528 GluLeuHisGlyTyrGlnGluLysPheHisArgProIleValMetThrGlyTyrGlyAla 547  
 DB 8009 GAACCTTGGCTGGCAGGAGAACTGCATCAGCGCATTAATCATCACGGAATACGGCGTG 8068  
 QY 548 AspThrIleuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGlnGluPheGlnVal 567  
 DB 8069 GATACGTTAGCGGAGCTGCATCATATGACACCGACATGTGGAGTGAAGATACAGTGT 8128  
 QY 568 GlnMetLeuAspMetTyrHisArgValPheAspArgIleGluSerMetAlaGlyGlnHis 587  
 DB 8129 GCATGGCTGAGTATATGATACACCGGCTTTGATGCGCTACGCCGCTGCGGTGAACAG 8188  
 QY 588 ValTrpAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnLys 607  
 DB 8189 GTATGGAATTTGCGCGATTTTGCACTCGCAAGGCAATATTGCGCGTGGCGGTACAG 8248  
 QY 608 LysGlyValPheThrArgAspArgLysProLysAlaAlaAlaHisSerLeuValGalaArg 627  
 DB 8249 AAAGGATCTTCACTCGGACCGCAACCGAAGTGGCGGCTTTTCTGTGCAAAAACGC 8308

QY 628 TrpTherSerIleasp 632  
 DB 8309 TGGACTGCGATGAC 8323

## RESULT 13

AAD09980/c  
 AAD09980 standard; DNA; 5919 BP.

XX AAD09980;

DT 12-SEP-2001 (first entry)

DE HSP-GUS construct for transfection of pacific oysters.

XX Fruitfly; fertility; reproduction; gametogenesis; microinjection; GUS;

KW glucuronidase; infection; plasmid pHSP-GUS; HSP; heat shock promoter;

XX embryogenesis; ds.

XX Drosophila melanogaster.

OS unidentified.

OS Chimeric.

PN WO200148224-A1.

PD 05-JUL-2001.

PF 22-DEC-2000; 2000NO-AU001596.

PR 24-DEC-1999; 99AU-00004884.

XX (CSIR ) COMMONWEALTH SCI & IND RES ORG.

PI Thresher R, Hinds L, Hardy C, Whyard S, Vignarajan S, Grewe PM;

XX Patil U;

XX WPI; 2001-425672/45.

XX Novel construct for preventing embryogenesis in animals comprises native

PT promoter, blocking DNA which abrogates function of crucial gene and

PT genetic switch to regulate expression/repression of blocker/gene

PT knockout.

PS Example 8; Page 196-198; 241pp; English.

XX The invention relates to a construct which allows animals to be bred in

CC captivity but renders them infertile in the wild by allowing reversible

CC control over fertility and reproduction. The construct comprises a native

CC promoter, a blocking DNA sequence contoured for and designed to abrogate

CC a crucial gene's function or to cause its mis-expression, and a genetic

CC switch to regulate controlled expression/repression of the blocker/gene

CC knockout. The construct is useful for preventing embryogenesis or

CC gametogenesis in animals by stably transforming an animal cell with the

CC construct by microinjection, transfection or infection, where the

CC construct stably integrates into the genome by homologous recombination,

CC and implanting the cell into a host organism, where a whole animal

CC develops from the implanted cell. The present DNA sequence is plasmid

CC construct pHSP-GUS construct used for transfection of Pacific oysters.

CC The plasmid contains glucuronidase (GUS) gene under the control of

CC Drosophila heat shock promoter (dHSP)

XX Sequence 5919 BP; 1504 A; 1474 C; 1414 G; 1527 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 2,49e-138

Score: 1714.00

Best Local Similarity: 70.50%

Query Match: 51.10%

DB: 4

Length: 5919

Matches: 331

Conservative: 92

Mismatches: 161

Indels: 16

Gaps: 9

US-10-757-093-4 (1-634) X AAD09980 (1-5919)

QY 40 ValArgProGlnArgTherSerArgGluLeuValAsnLeuAspGlyLeuTyrPhe 59  
 DB 2462 GTCCGTCTCTAGAAACCCCAACCGCTGAATCAAAAACTCGACCGCTGTGGCATTC 2403  
 QY 60 AlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTrrAlaProlu 76  
 DB 2402 AGCTGTGATCGCAAAACCTGTGGAAATGATCAG-----CGTGGTGGAAACCGCGTTA 2349  
 QY 77 ProlveGlyLeuGluCyProValProAlaSerTyrAsnAspIlePheIleSerArgLu 96  
 DB 2348 CAAGAAACCGGCAATTCGTGTGCAAGGCAAGTTTAAAGATAGTTCCGCATGCAGAT 2289  
 QY 97 IleHisAspHisValGlyTyrValTyrTyrGlnArgGluValIleValProlveGlyTyr 116  
 DB 2288 ATTCGTAAATTATTCGGGCAACGCTGTGTATCAGCGGAAGTCTTTATACGAAAGTTGG 2229  
 QY 117 SerGlnGluArgTyrLeuValArgAlaGluSerAlaThrHisIleGlyArgIleTyrVal 136  
 DB 2228 GCAGGCCAGCGTATCGCTGCTTCGATGCGGTCACTCATTAACGCAAAAGTGGGTC 2169  
 QY 137 AsnAspArgLeuValAlaGluHisValGlyTyrThrProPheGluAlaAspValThr 156  
 DB 2168 AATTAATCAGAAATGATGAGCATCAGGGGCTATACGCCATTTAAGCCGATGTACAG 2109  
 QY 157 GluLeuValAlaProGlyGluArgPheArgLeuThrIleGlyValAsnAsnGluLeuThr 176  
 DB 2108 CCGTATGTTATTCGGGAAAGTGTACGATACCGCTTGTGTAACAAGATCGAAC 2049  
 QY 177 HisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysArgIleGln 196  
 DB 2048 TGGCAGACTATCCCGCGGAAATGTGATTACC---GACCAAAACGGCAAGAAAGCAG 1992  
 QY 197 ThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIleTyrPheTyrSer 216  
 DB 1991 TCTTACTTCATGATTTCTTAACTAGCCGATACGCGATACGCGAATGCTTACACC 1932  
 QY 217 ValProGlnGlnHisIleGlnAspIleThrValIleThrAspVal-----AspGlyAsp 234  
 DB 1931 AGCCCGAACCCTGGGTGACGATACCGTGTACCGATCCGCGCAAGCTGTAAC 1872  
 QY 235 AsnGlyLeuLeuAsnTyrGluValAlaGluValAlaAsnGlnThrThrGlyGlnIleGlnIle 254  
 DB 1871 CACGCGTCTGTGATCGCAGGTG---GTGGCCAAAT-----GTTGATGTACAGGTT 1824  
 QY 255 SerValIleAspGluAspGlyAlaIleValAlaIleValAlaSerGlyAlaGlnGlyThrVal 274  
 DB 1823 GAACCTCGTATCGGATCAACAGGTGTGCAACGCAACGACCTAGCGGACCTTTG 1764  
 QY 275 ThrIleProSerValLysLeuTyrGlnProGlyAlaAlaTyrLeuTyrGlnLeuGlnVal 294  
 DB 1763 CAAGTGTGAATCCGACCTTGGCAACCGGTGAAGGTTATCTTATGAACCTGTGCTC 1704  
 QY 295 AsnIleValGlySerSerGlyAspValValAspThrTyrAsnLeuAlaThrGlyValArg 314  
 DB 1703 -----ACAGCCAAAGCCAGACAGATGATATACCCGCTTGGCGGCAATCCGG 1650  
 QY 315 ThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysProPheTyrPheThrGly 334  
 DB 1649 TCAGTGCAGATCGAAGGCGCAACGTTCTGATTAACACCAAAACGTTTACTTACCTGCGC 1590  
 QY 335 PheGlyLysHisGluAspThrAlaValArgGlyLysGlyHisAspProAlaTyrMetVal 354  
 DB 1589 TTGGTGTGATGAAATGGCGACCTTACGTTGCGCAAGATTCGATTAACGCTGATGGTG 1530  
 QY 355 HisAspPheGlnLeuMetLysTyrIleGlyAlaAsnSerPheArgThrSerHisTyrPro 374  
 DB 1529 CACGACACAGCATTAATGACGTGATTTGGGCGCAATCTCCACGTCACCTGATTAACCTT 1470  
 QY 375 TyrAlaGluGluValMetAspPheAlaAspArgAsnGlyIleValValIleAspGluThr 394  
 DB 1469 TACGCTGAAGAGATGCTCGATCGGCAAGATGAACATGCGATCTGTGTGATTTGAACCT 1410  
 QY 395 ProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGluSerGlyAlaPro--- 412

```

Db      1409 GCGTGTGCGGCTTTAACTCTTTAGGATGTTTGAAGCGGCAACAGCCGAA 1350
Qy      413 GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAlaHisLysGlnAlaIle 432
Db      1349 GAACCTGACCGCAAGGAGAGTCAACGGGAAACCTCAGAACCGACTTACCGGGAAT 1290
Qy      433 ArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMetTrpSerIleAlaAsn 452
Db      1289 AAAGACTGATGACCGGTGACAAAACCAACCAAGCTGTGTATGTGGAGTATGGCAAC 1230
Qy      453 GluProAlaSerHisGlnLysGluValArgGluTyrPheGluProLeuThrAsnLeuThr 472
Db      1229 GAACCGGATACCCGTCGCAAGGTGACCGGATATTTCCGCGCACTGGCGGAACGACG 1170
Qy      473 ArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyrGln 492
Db      1169 CGTAACTGACCGGCGGTCCGATCACCCTGCGTCAATGTATGTTTCCGACGCTCAC 1110
Qy      493 LeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTyr 512
Db      1109 ACCGATACCATCGCATCTCTTGTATGTCTGTCTGTACACCGTTATTAAGATGTAT 1050
Qy      513 SerGlnThrGlyAspLeuGluGluAlaGluAlaIleLeuGluLysGluLeuHisGlyTyr 532
Db      1049 GTCCAAAGCGCGCATTTGGAAACGACGAGAAAGTACTGGAAAAAGAACTTGGCCTCG 990
Qy      533 GlnGluLysPheHisArgProIleValMetThrGluTyrGlyAlaAspThrLeuAlaGly 552
Db      989 CAGGAAACCTGATCAGCCGATTCATCACCGAAATACCGCGCTGTAGTACGCGG 930
Qy      553 LeuHisSerIleLeuGlyLeuProTyrSerGluGluPheGlnValGlnMetLeuAspMet 572
Db      929 CTGCATCTCATATGACACCGACATGTGAGTGAAGAGATCATAGTGCATGCGTGAATAG 870
Qy      573 TyrHisArgValPheAspArgIleGluSerMetAlaGlyLysIleValTyrAsnPheAla 592
Db      869 TATCACCGCGCTTTGATCGCTCAGCGCGCTGCTGCTGCAACAGATATGAAATTCGCC 810
Qy      593 AspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnLysGlyValPheThr 612
Db      809 GATTTTGGCAGCTTCGACGAGCATATTTGGCGGTGGCAAGAAAGGATCTTCACCT 750
Qy      613 ArgAspArgLysProLysAlaIleAlaHisSerLeuArgLysIleThrTrpTrpSerIleAsp 632
Db      749 CCGGACCGCAACCGAAGTGGCGGCTTTCTGCTGCAAAAACGCTGAGCTGGCATGAC 690

RESULT 14
ADM01287
ID      ADM01287 standard; DNA; 7510 BP.
XX
AC      ADM01287;
XX
DT      01-JUL-2004 (first entry)
XX
DE      Plasmid pUTPA-f-GUS.
XX
KW      vector; plasmid; artificial intergene region; plant;
KW      transplastomic angiosperm; agronomic property; stress resistance;
KW      tbc1 gene; ds.
XX
OS      Synthetic.
XX
PN      WO2004029256-A2.
XX
PD      08-APR-2004.
XX
PF      15-SEP-2003; 2003WO-CU000009.
XX
PR      27-SEP-2002; 2002CU-00000208.
XX
PA      (ING-) CENT ING GENETICA & BIOTECNOLOGIA.
XX

```

```

PI      Selman-Housein Soea G, Aguilar Cabeza E, Gonzalez Quintero ADC;
PI      Ramos Gonzalez O;
XX
DR      MPI, 2004-316131/29.
XX
PT      DNA vector for transformation and expression in plasmids, useful e.g. for
PT      producing pharmaceutical proteins or improving agronomic properties, has
PT      gene inserted in artificial intergene region.
XX
PS      Example 2; SEQ ID NO 19; 74pp; Spanish.
XX
CC      The invention relates to a DNA vector (A) for stable transformation and
CC      expression of genes (I) in plasmids, where (I) is inserted in an
CC      artificial intergene region (AIR) formed by combining two 5'-untranslated
CC      regions (5'-UTRs) of genes that transcribe in different directions and
CC      are derived from plants of different divisions or classes. (A) are used
CC      to produce transplastomic angiosperms that have improved agronomic
CC      properties (e.g. resistance to biotic or abiotic stress) or express a
CC      very wide range of agricultural, veterinary, pharmaceutical, nutritional
CC      or industrial products, e.g. enzymes, vaccinating antigens, cytokines or
CC      immunoglobulins. Use of (A) eliminates the need for a transposon for gene
CC      insertion; inserted genes do not require promoters and terminators; and
CC      the structure of flanking sequences in (A) ensures universal
CC      applicability. Also any selection marker in (A) can be eliminated by
CC      homologous recombination. (A) provides efficient and stable expression of
CC      genes without causing any functional alterations. This sequence
CC      represents the plasmid pUTPA-f-GUS an example of the vector of the
CC      invention.
XX
SQ      Sequence 7510 BP; 2025 A; 1695 C; 1867 G; 1923 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      3,39e-138      Length:      7510
Score:          1714.00      Matches:      331
Percent Similarity: 70.50%      Conservative: 92
Best Local Similarity: 55.17%      Mismatches: 161
Query Match:     51.10%      Indels:      16
DB:              12      Gaps:      9

US-10-757-093-4 (1-634) x ADM01287 (1-7510)
Qy      40 ValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTyrLysPhe 59
Db      1924 GTACGCTCGTGTAAACCCCAACCCCGTGAACCAAAAACCTGCACCGCGCTGGGCGATTC 1983
Qy      60 AlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProThrPheLysPheLeu 76
Db      1984 AGCTGTGATCGCAAAACCTGTGAATTGATCAG-----CGTTGGTGGAAAGCGGCTTA 2037
Qy      77 ProLysGlyLeuGluCysProValProAlaSerTyrAsnAspIlePheIleSerArgGlu 96
Db      2038 CAAGAAAGCGGGCAATTCCTGTGCGCAGGACAGTTTAAAGATCAGTTCGCCGATCGAGAT 2097
Qy      97 IleHisAspHisValGlyTyrValTyrTyrGlnArgGluValIleValProLysGlyTyr 116
Db      2098 ATTCTGAATTAATGCGGCAACGCTGTGTATCAGCGGCAAGTCTTTATACCGAAAGGCTTG 2157
Qy      117 SerGlnGluArgTyrLeuValArgAlaGluSerAlaThrHisIleAspGlyArgIleTyrVal 136
Db      2158 GCAAGCCACGATCTGCTGCTGGCGTTTCGATGGCGTCATCATTAACGCAAGGTGGGTC 2217
Qy      137 AsnAsnArgLeuValAlaGluHisValGlyGlyTyrThrProPheGlnAlaAspValThr 156
Db      2218 AATTAATCAGAGAGTAGTAGAGATCAGGCGGCTATACGCCATTTGAACCCGATGTCACG 2277
Qy      157 GluLeuValAlaProGlyLysPheArgLeuThrIleGlyValAsnAsnGluLeuThr 176
Db      2278 CCGTATGTATTGCCCGGAAAGTGTATACCGTATCCGTTGTGTGAACAACGAACCTGAC 2337
Qy      177 HisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysArgIleGln 196
Db      2338 TGGCAGACTATCCCGCGGGAATGTGATTTACC--GACGAAAAACGCAAAAAAGCAG 2394

```

QY	197	ThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaGlySerIleTyrPheLeuSer	216
Db	2395	TCTTACTTCACAGATATTTCTTTAACTATGCGGAAATCCAGATCCAGATATGCTTCAACC	2454
QY	217	ValProGlnGlnHisIleGlnAspIleThrValValThrAspVal-----AspGlyAsp	234
Db	2455	ACGCCGAACACCTGGGTGGACGATATCACTGGTGGACGATGTGCGCAAGACTGTATAC	2514
QY	235	AsnGlyLeuIleAsnTyrGlnValGlnValAlaAsnGlnThrGlyGlnIleGlnIle	254
Db	2515	CACCCGCTGTGGTCACTGGCAGGTG---GTGGCCAAAT-----GGTGAATGCAGCGTT	2562
QY	255	SerValIleAspGluAspGlyAlaIleValAlaIleAsnIleSerGlyAlaGlnGlyThrVal	274
Db	2563	GAACCTCGTGAATGGCGATCAACAGGTGGTGTCAACTGGACAAAGCACTAGCGGGACTTGG	2622
QY	275	ThrIleProSerValIleLeuTyrGlnProGlnProGlnAlaIleTyrLeuTyrGlnLeuGlnVal	294
Db	2623	CAAGTGGGAATCCGCACCTCTGGCAACCGGGTGAAGTTATCTCTAAGAACTGTGCGTC	2682
QY	295	AsnIleValGlySerSerGlyAspValValAspThrTyrAsnLeuAlaThrGlyValArg	314
Db	2683	-----ACAGCCAAAAGCCAGACAGAGTGAATCTACCCGCTTCGCGTGGCATCCGG	2736
QY	315	ThrValIleValAlaGlySerGlnPheLeuIleAsnGlyLeuProPheThrPheThrGly	334
Db	2737	TCAATGGCAGAGTGAAGGGGCAACAGTCTCTGATTAACCAAAACCGTTTACTTATACGGC	2796
QY	335	PheGlyLeuSerIleGluAspThrAlaValArgGlyLeuGlyIleHisProAlaTyrMetVal	354
Db	2797	TTTGGTCTCATGAAGATCGCGACTGGCTGGCAAAAGATTCGATACGTGCTGATGGTG	2856
QY	355	HisAspPheGlnLeuMetLeuTyrIleGlyAlaAsnSerPheArgThrSerHisTyrPro	374
Db	2857	CACGACCAACGATTAATAGACTGGATGGGGCAACTCTCAACCGTACTCGCATTAACCT	2916
QY	375	TyrAlaGlnGluValMetAspPheAlaAspArgAsnGlyIleValValIleAspGluThr	394
Db	2917	TACCGTGAAGAGATCTCTGCACTGGCGCAATGAACATGGCAATCGTGATTTGATGAATAACT	2976
QY	395	ProAlaValIleGlyLeuAsnIleAlaLeu---MetGlyValSerGluSerGlyAlaPro---	412
Db	2977	GCTGCTGTCGGCTTTAACTCTCTTTTAAAGCATGGTTTTCGAAGCGGCAACAGCCGAAA	3036
QY	413	GlnThrPheThrProAspAlaIleAsnAspIleThrGlnGlnAlaHisIleGlnAlaIle	432
Db	3037	GAACGTGACAGGAAGAGGACAGTCAACGGGGAAATCAGCAAGCCCACTTACAGCGCATT	3096
QY	433	ArgGluLeuIleAlaIleAspIleAsnHisAlaSerValIleMetTyrSerIleAlaAsn	452
Db	3097	AAAGAGCTGAATGACCGCTGACAAAACCAACCCAGCCGTGGATGTGAGATTTGCCAAC	3156
QY	453	GluProAlaSerHisIleGluAspGlyAlaIleArgGluTyrPheGluProLeuThrAsnLeuThr	472
Db	3157	GAACCGGATACCCGCTCCGCAAGGTGCAGGGGAATTTTCGGCCACTGGCGGAAACAG	3216
QY	473	ArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyrGln	492
Db	3217	CGTAAACTCGAACCCGACCGCTGCACTACCTGCGCAATGTATGTTGCGGACGCTCAC	3276
QY	493	LeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTyrTyr	512
Db	3277	ACCGATACCAATCAGCATCTCTTGATGTGCTGCTGTCGTAACCCGTTATTCAGAGGTAT	3336
QY	513	SerGlnThrGlyAspLeuGlnGluAlaGluAlaIleAsnGluTyrGlnLeuHisIleGlyTyr	532
Db	3337	GTCGAACGGGGGATTTGGAAACGGGAGAAAGTATCTGGAAAAGAACTTCTGCGCTGG	3396
QY	533	GlnGlnIlePheHisIleArgProIleValMetThrGluTyrGlyAlaAspThrLeuAlaGly	552
Db	3397	CAGAGAAATGCAATCAGCGCATTTATCATCAACGAAATACGGGTGGATATCGTTACCGGG	3456
QY	553	LeuHisSerIleLeuGlyLeuProTyrSerGluGluPheGlnValGlnMetLeuAspMet	572

Db 3457 CTGACACTCAATNGTACACGCACATCTGGAGTGAAGAAGATACGTGTGCATGGCTGCATATG 3516  
 Oy 573 TyrHisArgValPheAspArgIleGluSerMetAlaGlyIuHisValTTPAsnPheAla 592  
 Db 3517 TATACACCGCGCTTGTGATCGCGTCAGCGCCCTCGCTCGGTGAACAGATATGAAATTTCGCC 3576  
 Oy 593 AspHegInThrAnLeuEngIylleIlaArgValAspGlyAsnIyIySGIYValPheThr 612  
 Db 3577 GATTTTTCGACCTTCGCAGAGCATATTCGGCTTCGGCTTAACAAGAAAGGATCTTCACT 3636  
 Oy 613 ArgAspArgIySPolYSAlaAlaAlaHisSerLeuArgAlaArgTTPHSerIleAsp 632  
 Db 3637 CGGACCGCCAAACCGAAGTCTCGCGCTTTTCTGTCGCAAAAACCTGACCTGGCATGAAC 3696  
 RESULT 15  
 ID ADM01291  
 XX ADM01291 standard; DNA; 8327 BP.  
 AC ADM01291;  
 XX 01-JUN-2004 (first entry)  
 DE Plasmid pVTPA-f-GUS-aada DNA sequence.  
 XX vector; plasmid; artificial intergene region; plant;  
 KM transplastomic angiosperm; agronomic property; stress resistance;  
 KM rbcL gene; ds.  
 XX  
 OS Unidentified.  
 PN WO2004029256-A2.  
 XX 08-APR-2004.  
 PD 15-SEP-2003; 2003WO-CU000009.  
 PF 27-SEP-2002; 2002CU-00000208.  
 PR (INGG-) CENT ING GENETICA & BIOTECNOLOGIA.  
 PA Selman-Housein Sosa G, Aguilar Cabeza E, Gonzalez Quintero ADC;  
 XX Ramos Gonzalez O;  
 PI WPI; 2004-316131/29.  
 DR  
 XX  
 PS Example 4; SEQ ID NO 23; 74pp; Spanish.  
 CC The invention relates to a DNA vector (A) for stable transformation and  
 CC expression of genes (I) in plasmids, where (I) is inserted in an  
 CC artificial intergene region (AIR) formed by combining two 5'-untranslated  
 CC regions (5'-UTRs) of genes that transcribe in different directions and  
 CC are derived from plants of different divisions or classes. (A) are used  
 CC to produce transplastomic angiosperms that have improved agronomic  
 CC properties (e.g. resistance to biotic or abiotic stress) or express a  
 CC very wide range of agricultural, veterinary, pharmaceutical, nutritional  
 CC or industrial products, e.g. enzymes, vaccinating antigens, cytokines or  
 CC immunoglobulins. Use of (A) eliminates the need for a transposon for gene  
 CC insertion; inserted genes do not require promoters and terminators; and  
 CC the structure of flanking sequences in (A) ensures universal  
 CC applicability. Also any selection marker in (A) can be eliminated by  
 CC homologous recombination. (A) provides efficient and stable expression of  
 CC genes without causing any functional alterations. This sequence  
 CC represents the plasmid pVTPA-f-GUS-aada, an example of the vector of the  
 CC invention.  
 CC  
 CC Sequence 8327 BP; 2228 A; 1894 C; 2098 G; 2107 T; 0 U; 0 Other;  
 SQ Alignment Scores:

Pred. No.: 3,87e-138 Length: 8327  
 Score: 1714.00 Matches: 331  
 Percent Similarity: 70.50% Conservative: 92  
 Best Local Similarity: 55.17% Mismatches: 161  
 Query Match: 51.10% Indels: 16  
 DB: 12 Gaps: 9

US-10-757-093-4 (1-634) x ADM01291 (1-8327)

QY 40 ValArgProGlnrghThrsSerSerArgGluLeuValAsnLeuaspGlyLeuTyrPhe 59  
 DB 1924 GlnAGTCCTGTAGAAACCCACACCGGATGAAATCAAAAACCTCAACGCGCTGTGGCAATTC 1983  
 QY 60 AlaLeu-----AlaSerGlyLeuAsnaspThrLaglnProTPrLThrLProLeu 76  
 DB 1984 AGTCGTGATCGGAAACCTGTGATTCAG-----CGTTGGTGGGAAACCGGCTTA 2037  
 QY 77 ProlyGlyLeuGlnLucyProValProAlaSerTyrAsnaspLlePheLleSerArgGlu 96  
 DB 2038 CAAGAAAGCGGCAATTGCTGTGCCAGCGATTTTAAACATCATCATGTCAGAT 2097  
 QY 97 IleHisAspHisValGlyTyrPheValTyrTyrGlnArgGluValLleValProlyGlyTyr 116  
 DB 2098 ATTCGTAATTAATCGGGCAACGTCGTGATCACCGCAAGCTCTTTATACGAAAGGTGG 2157  
 QY 117 SerGlnGluArgTyrLeuValArgAlaGluSerAlaThrHisIsglyArgLleTyrVal 136  
 DB 2158 GCAGGCGAGGTATCGTGTGCGCTTCGATGCGGTCACTCATACGCAAGGTGGTTC 2217  
 QY 137 AsnAsnArgLeuValAlaGluHisValGlyTyrTyrProPheGluAlaAspValThr 156  
 DB 2218 AATTAATCAGGAGTATGAGCATCAGGCGGCTTATACGCTTGAAGCGCATGTCCAG 2277  
 QY 157 GluLeuValAlaArgGlyGluLysPheArgLeuThrLleGlyValAsnGlnLeuThr 176  
 DB 2278 CCGTATGTTATTCGCCGAAAGTGAATGATACCGGTTGTGTGAACAACGAACTGAAC 2337  
 QY 177 HisGluThrLleProProGlyLysLleThrThrGlyAsnAlaThrGlyLysArgLleGln 196  
 DB 2338 TGGCAGCATATCCGCGCGGAGATGTGATTAAC---GACGAAACCGCAAGAAAGACGAG 2394  
 QY 197 ThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerLleTyrPheTyr 216  
 DB 2395 TCTTACTTCATATTTCTTAACTATGCGGAGATCATCGACGCTTATGCTCTACACC 2454  
 QY 217 ValProGlnGlnHisLleGlnAspLleThrValLleThrAspVal-----AspGlyAsp 234  
 DB 2455 ACGCCGAACACCTGGGTGACGATATCACCGGTGACGATGTCGCGCAAGCTGTATC 2514  
 QY 235 AsnGlyLeuLleAsnTyrGluValAlaGluAlaAsnGlnThrThrGlyGlnLleGlnLle 254  
 DB 2515 CACGCGTCTGTTGACGTGCAGGTG---GTGGCCAAAT-----GGTATGTCAACGTT 2562  
 QY 255 SerValLleAspGlnAspGlyAlaLleValAlaLysAlaSerGlyAlaGlnGlyThrVal 274  
 DB 2563 GAACCTCGTATCGGATCAACAGGTGTGCACTGCAACAGGCACTAGCGGACCTTGG 2622  
 QY 275 ThrLleProSerValLysLeuTyrGlnProGlyAlaLleValTyrLeuTyrGlnLeuGlnVal 294  
 DB 2623 CAAGTGTGATCCGACCTCTGGCAACGGGTGAAGGTATCTCATGAACGTGTCCGTC 2682  
 QY 295 AsnLleValGlySerSerGlyAspValValAspThrTyrAsnLeuAlaThrGlyValArg 314  
 DB 2683 -----ACAGCCAAAGCCAGACAGAGTGTGATATCTACCCGCTTCGCGGCATCCGG 2736  
 QY 315 ThrValLysValAlaGlySerGlnPheLeuLleAsnGlyLysProPheTyrPheThrGly 334  
 DB 2737 TCAGTGTGAGTGAAGGCGCAACAGTTCGTATTAACCAACAACCGTCTTACTTATGAGC 2796  
 QY 335 PheGlyLysHisGluAspThrAlaValArgGlyLysGlyHisAspProAlaTyrMetVal 354  
 DB 2797 TTGTGTGTATGAATGCGGACCTTGTGCGCAAGAGATTCGATTAACGTGTGATGTG 2856

QY 355 HisAspPheGlnLeuMetLysTyrLleGlyAlaAsnSerPheArgThrSerHisTyrPro 374  
 DB 2857 CACGACCAAGCATTTATGACATGGATTGGGGCAATCTCTACCGTACCTCGCATTCCT 2916  
 QY 375 TyrAlaGlnGluValMetAspPheAlaAspArgAsnGlyTyrLleValAlaLleAspGluThr 394  
 DB 2917 TACGCTGAAGAGATGCTCCACTGGGACGATGAACATGATCATGTGTGATGTAAGT 2976  
 QY 395 ProAlaValGlyLeuAsnLleAlaLeu---MetGlyValSerGluSerGlyAlaPro--- 412  
 DB 2977 GCTGTGTGGCTTTTAACTCTCTTTAGGCATTTGTTGAAAGCGGCAACAGCGGAAA 3036  
 QY 413 GlnThrPheThrProAspAlaLleAsnAspLysThrGlnGluAlaHisIsglyGlnAlaLle 432  
 DB 3037 GAACGTACAGCGAAGAGCAGTCAACCGGAAACCTCAACAGCGCATTTACAGCGGAT 3096  
 QY 433 ArgGluLeuLleAlaArgAspLysAsnHisAlaSerValValMetTyrSerLleAlaAsn 452  
 DB 3097 AAGAGCTGATAGCGGTGACAAAACCAACCAAGCGTGTGATGTGAGTATTCGCAAC 3156  
 QY 453 GluProAlaSerHisGluAspGlyAlaArgGluTyrPheGluProLeuThrAsnLeuThr 472  
 DB 3157 GAACCGGATACCGGTCCGCAAGGTGACGGAAATATTTCCGCACTGGCGGAAAGCAAG 3216  
 QY 473 ArgGlnLeuAspProThrArgProLleThrPheAlaAsnValGlyThrAlaThrTyrGln 492  
 DB 3217 CGTAAACCTGACCCGACGCGGTCCGATCACCTCGTCAATGATGTTCGCAAGCTCAC 3276  
 QY 493 LeuAspArgLleSerAspLeuPheAspValSerCysLleAsnArgTyrPheGlyTyrPhe 512  
 DB 3277 ACCGATACCATCAGCATCTCTTGTGTGCTGTGTGACCGGTATATACGATGTAT 3336  
 QY 513 SerGlnThrGlyAspLeuGlnGluAlaGluAlaLleGluLysGluLeuHisIsglyTyr 532  
 DB 3337 GTCCAAAGCGGCGATTGGAACGCGCAAGATCTGGAAGAAAGAACTTCGCGCTGG 3396  
 QY 533 GlnGluLysPheHisArgProLleValMetThrGluTyrGlyAlaAspThrLeuAlaGly 552  
 DB 3397 CAGGAGAACTCATCAGCGCATTTATCATCACGGAATACGGCGTGTGATGCTTACGCGGG 3456  
 QY 553 LeuHisSerLleLeuGlyLeuProTyrPheSerGlnGluPheGlnValGlnMetLeuAspMet 572  
 DB 3457 CTGCATCATATGTACACGACATGTGAGTGAAGATCAAGTGTGATGCTGTGATATG 3516  
 QY 573 TyrHisArgValPheAspArgLleGlnSerMetAlaGlyGlnHisIsglyThrAspPheAla 592  
 DB 3517 TATCACCGGCTCTTGTGATGCGTCAAGCGCGTGTGCGTGAACAGGTATGAAATTCGCC 3576  
 QY 593 AspPheGlnThrAsnLeuGlyLleLleArgValAspGlyAsnLysGlyValAlaPheThr 612  
 DB 3577 GATTTTGCACCTCGCAAGCATATTTGGCGGTGATCAAGAAAGGATCTTCACT 3636  
 QY 613 ArgAspArgLysProLysAlaAlaHisSerLeuAspGlyAlaGlyTyrThrSerLleAsp 632  
 DB 3637 CCGGACCGCAACCGAAGTGGCGGCTTTTGTGTCGCAAAAACGCTGTGATGTCATGAC 3696

Search completed: March 22, 2005, 18:38:23  
 Job time : 885 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 22, 2005, 16:22:20 ; Search time 6389 Seconds

(without alignments)  
4808.359 Million cell updates/sec

Title: US-10-757-093-4  
Perfect score: 3354  
Sequence: 1 MKFLGLSLSLAPSLGNP.....RKPKAAHSLRARMTSIDKN 634

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 segs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seg length: 0  
Maximum DB seg length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame\_plus\_p2n.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool.p/US10757093/runat\_18032005\_164455\_27764/app\_query.faeta.1.775  
-DB=GenEmbl -Qfmt=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10757093 @CGN\_1\_1\_5600@runat\_18032005\_164455\_27764 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLIG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_hcg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pac: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_scs: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

No.	Score	Query Match	Length	DB ID	Description
1	1722	51.3	7599	12	EVPN0102
2	1721.5	51.3	3169	6	A57359 Sequence 5
3	1721.5	51.3	3169	6	AR096536 Sequence
4	1721.5	51.3	3169	6	AR373248 Sequence

c	5	1718.5	51.2	9037	12	AY286001	Cloning v
c	6	1718.5	51.2	10313	12	AY237647	Cloning v
c	7	1718.5	51.2	11973	12	AY237648	Cloning v
c	8	1718	51.2	2103	1	AF305918	Escherich
	9	1717	51.2	4299	6	AK025617	Sequence
	10	1716	51.2	4299	6	ATK298139	Arabidops
	11	1716	51.2	11509	8	ATK298137	Arabidops
	12	1715.5	51.1	5667	12	AF502128	Transfient
	13	1715.5	51.1	5667	12	EC041412	Transfient
	14	1715.5	51.1	14701	12	AY456412	Cloning v
	15	1715.5	51.1	14758	12	AY485783	Cloning v
	16	1715	51.1	9152	12	AF527485	Cloning v
	17	1714.5	51.1	8654	6	AX093006	Chloroplast
	18	1714.5	51.1	8654	6	AX093009	Sequence
	19	1714.5	51.1	8987	6	AX093017	Sequence
	20	1714.5	51.1	9390	6	AX093013	Sequence
	21	1714.5	51.1	9390	6	AX093015	Sequence
	22	1714.5	51.1	9688	6	AX093045	Sequence
c	23	1714.5	51.1	15208	6	AX093046	Sequence
	24	1714	51.1	4519	12	U02456	Sequence
c	25	1714	51.1	6875	12	AF433042	Cloning v
c	26	1714	51.1	6828	12	AF433043	Cloning v
c	27	1714	51.1	7510	6	CQ797813	Sequence
c	28	1714	51.1	8327	6	CQ797817	Sequence
c	29	1714	51.1	32798	6	AR343138	Sequence
c	30	1714	51.1	32798	6	AX382187	Sequence
c	31	1713.5	51.1	4995	12	XXU02841	Cloning v
c	32	1712.5	51.1	3035	6	AR030540	Sequence
c	33	1712.5	51.1	19211	1	D90805	Sequence
c	34	1712.5	51.1	110000	1	U00096.16	Sequence
c	35	1712	51.0	8367	12	AY562545	Sequence
c	36	1712	51.0	9285	6	AX093047	Sequence
c	37	1712	51.0	9342	12	AY562544	Sequence
c	38	1712	51.0	11226	12	AF234306	Binary ve
c	39	1712	51.0	11226	12	AF234312	Binary ve
c	40	1712	51.0	11378	12	AF234294	Binary ve
c	41	1712	51.0	11378	12	AF234295	Binary ve
c	42	1712	51.0	11633	12	AF234316	Binary ve
c	43	1712	51.0	11785	12	AF234314	Binary ve
c	44	1712	51.0	11849	12	AF234297	Binary ve
c	45	1712	51.0	12001	12	AF234293	Binary ve

#### ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	REMARK	REFERENCE	AUTHORS
EVPN0102	LOCUS	Expression vector (PNOM102) DNA, 7599bp.	EVPN0102	232701.1	GI:475168	beta-glucuronidase; gpda gene; uidA gene.	synthetic construct	other sequences; artificial sequences.	Mullane, E.J., Hamer, J.E., Roberti, K.A., Yelton, M.M. and Timberlake, W.E.	Primary structure of the trpC gene from <i>Aspergillus nidulans</i>	Mol. Gen. Genet. 199 (1), 37-45 (1985).	85213104	3158796	(sites)	2	Roberts, I.N., Oliver, R.P., Punt, P.J. and van den Hondel, C.A.
																Expression of the <i>Escherichia coli</i> beta-glucuronidase gene in industrial and phytopathogenic filamentous fungi
																Curr. Genet. 15 (3), 177-180 (1989)
																89354666
																2504501
																(sites)
																3 (bases 999 to 2572)
																Punt, P.J., Dingemans, M.A., Kuyvenhoven, A., Soede, R.D.,

```

misc_difference 3136 /replace="a"
/gene="uidA"
/note="copy mistake in the genbank entry; see Jefferson ET
Al., 1986, PNAS 83:8447-8451"
/citation=[2]
/replace="g"
/misc_difference 3560. 3579
/gene="uidA"
/note="sequence corrections as in Farrell and Beachy
(1990) Plant Mol Biol 15:821-825"
/citation=[2]
/replace="gtgcacggaattccgcg"
4182. 4951
/standard name="terminator region from A.nidulans trpC
gene (containing CDS)"
/note="this region contains the 3'end of the trpC
transcript including 50 codons. two major transcription
ends have been identified"
/citation=[2]
/evidence=experimental
4433^4443
polyA site 4443. 4453
misc_difference 4443. 4453
/note="introduction of one base at this position generates
an BglI site with was shown to be present by restriction
digestion"
/citation=[1]
/replace="gccttcaggc"
4652. 4660
misc_difference 4652. 4660
/note="introduction of one base at this position generates
an AluNI site which was shown to be present by restriction
digestion"
/citation=[1]
/replace="cagcnctg"
4706^4716
polyA site 4706. 4750
misc_difference 4740. 4750
/note="introduction (actual sequence could be deletion as
well) of a base at this position results in the
elimination of an xmnI site which was shown to be absent
by restriction digestion"
/citation=[1]
/replace="gaatcanttc"
4777. 4784
variation 4777. 4784
/note="A NcoI site present at this position was removed by
restriction, treatment with Klenow polymerase and
religation"
/citation=[2]
/replace="catgcatg"
4952. 7599
misc_feature 4952. 7599
/product="pUC18"
/note="Vector sequences derived from pUC18 from SALI to
EcoRI"
/citation=[2]
ORIGIN
Alignment Scores:
Pred. No.: 6,3e-116 Length: 7599
Score: 1722.00 Matches: 335
Percent Similarity: 69.18% Conservative: 96
Best Local Similarity: 53.77% Mismatches: 172
Query Match: 12.34% Indels: 20
DB: 12 Gaps: 10
US-10-757-093-4 (1-634) x EVPMO102 (1-7599)
OY 21 AAlAlArGHisPheProArGaNgLuWeThrGlnHisGluGlnPro----- 36
Db 2236 TCAGTTCGAGCCTTCCCACTTCATCGACGACTTGACTACAGACTACCCGCTTGAGCAGAC 2295
OY 37 LeuIlelyValArGProGlnArGhrHeSerArGgIuLeValAsnLeuAspGlyLeu 56
Db 2296 ATACCATGATCGCTCTGTAGAAACCCCAACCCGCTGAAATCAAAAACTGACGGCCTG 2355

```





```

/strain="FRAJ210"
/db_xref="taxon:562"
/clone="FTRC99 DICISTR. FAB/E.C.-BETA-GLUC"
3..644
/note="unnamed protein product"
/codon_start=1
/translation="MDIQMTQSPSSLSASVGRVITTCSTSSVSVMHWYQKPKAP
KLLIYSTNLASGVPSRFSGSGSDFTFTISLQPEDATATYCHQMSYPRFGGKTK
VRIKRTVAASVFIPEPSDEOLKSGTASVCLNPNPBEAKVQKMDNALDQSGNSQ
SVTEQDSKDTSTLSLTLSKADYKHKVVAACEVTHQSLSPVTKFNRGRC"
666..3164
/note="unnamed protein product"
/codon_start=1
/translation="MEVQIQESGPGIVRPSQTLSTCTVSGFTTSSGYSWMHWQRPQ
RLEWIGYIQYSGITNPNPSLKSRTVMTVDISNOPSLSSTVTAADTVVYICABEDY
DTHWYFDVWGQSLVTVTVSSATKGPVPLAPCSRSTSGTALAGCLVKQYFPEPV
TVSNMGSALTSGVHTFPAVLQSSGLYSLSSVTVSSSLGTQYTCNVNMRKSNKVD
KVEILGSGSGSWVRPVEPTREIKKLGLMAFSLDENCGIDQRMWESALQESRAIAY
PESPNDPADADIRNAGNVMYQREVFIRPKMAGORIVRPAVTHDGVNNOEVM
EHQGGTPPEADVTPIVIAKSRITIVCNNEINMQITPRGAVITDENGKKKQSYFHN
FNYAGIHSVMLYTPNTWVDITVTHVADQCNHASVDKQVANGDVSRLRBDQ
QVATIQGTSGLQVNPMLQWQEGYELVELCVTAQSQTECDIYPLRIGIRSVAVAGE
OPLINRPFYFTGFRHEDADLRKGFMDLWMDHADMWIGANSYRISHYVPAEEM
LPMADHGIIVIDEITAGVFNLSLIGFPAKNPKELVEBEZANVGTQOALQAIKEL
IARDKNPVSVMISIANEPDTPROGAREAPLAELATRLDPRPITCNVWFCDAHT
PTISDLPTLCLNRYGMYVQSGDLTAEKAVLEKLLAOKELHOP11ITEYGVDTLA
GHSMTDMSEBYOCAMLDMHRVDFDRSAVGEVWMPFADPATSGILIRGANKKS
LETRDKRPSAFLQKRWTKMNFKEPQGGKQ"

```

## ORIGIN

## Alignment Scores:

```

pred. No.:      2 2e-116      length:      3169
Score:          1721.50      Matches:      339
Percent Similarity: 68.17%      Conservative: 98
Best Local Similarity: 52.89%      Mismatches: 179
Query Match:      51.33%      Indels:      25
                        Gaps:      11

```

US-10-757-093-4 (1-634) x A57359 (1-3169)

```

QY      8 SerLeuLeuSerLeuAlaAlaProSerLeuGlyThrProAla-----AlaArg 23
Db      1233 AGCGTGGTGAACCGCTCCAGAGCTTGCGGCAACCACTCAACCTGCAACGTAAT 1292
QY      24 HisPheProArgAnGluMetThrGlnHisGluGlnProLeu-----Ile 38
Db      1293 CACAAAGCCGACGACACACAAAGGTGACAAAGAGTTCGCTTCGTTCC 1352
QY      39 LysValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpLys 58
Db      1353 ATGGTACCTCTCTGAGAAACCCCAACCCCTGGAATCAAAAACTCCAGCGCCCTGTGGCA 1412
QY      59 PheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTrpThrAlaPro 75
Db      1413 TTCAAGTCTGATCGCAAAACCTGTGGAATTCATGAG-----CGTTGTGGAAAGCGCG 1466
QY      76 LeuProLysGlyLeuGluCysProValProAlaSerTyrAsnAspIlePheIleSerArg 95
Db      1467 TTCAAAAGAAAGCCGGAATTCCTGTCACGAGCGATTTCACATCACTTCGCGCATGCA 1526
QY      96 GluIleHisAspHisValGlyTrpValArgTrpGlnArgGluValIleValProLysGly 115
Db      1527 GATATTCTGTAATTATGCGGGCAACGTCGTGATACGCGCAAGCTTTATACGAAAGGT 1586
QY      116 TrpSerGlnGluArgTrpLeuValArgAlaGluSerAlaThrHisGlyValGlyLeuTrp 135
Db      1587 TGGGCAAGCGACGATATGCTGCTGTTTCGATGCGGCTCACTCATTCAGCGCAAAAGTGTGG 1646

```

```

QY      136 ValAsnAspArgLeuValAlaGluHisValGlyTrpThrProPheGluAlaAspVal 155
Db      1647 GTCAATTAATCAGAAAGTATGATGAGCATCAGGCGCGCTATACGCACTTTGAAGCCGATTC 1706
QY      156 ThrGluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyValAsnAsnGluLeu 175
Db      1707 AGCGCGTATGTTATGCGCGGAAAGTAGTGATACCGCTTGTGTGAAACCAACACTG 1766
QY      176 ThrHisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysArgIle 195
Db      1767 AACTCGCAGACTATCCCGCGGAAATGGTATAC--GACGAAACGGCAAGAAAAG 1823
QY      196 GlnThrTrpGlnHisAspPheTyrAsnTrpAlaGlyLeuAlaArgSerIleTrpLeuTrp 215
Db      1824 CAGTCTTACTTCATTAATTTCTTTAATCATTCGCGGATTCATCGAGCGTATGCTTAC 1883
QY      216 SerValProGlnGlnHisIleGlnAspIleThrValValThrAspVal-----AspGly 233
Db      1884 ACCACGCGCAACACTGGGTGGAGCATATCACCGGTGGAGCATGTCGCGCAACTGT 1943
QY      234 AspAsnGlyLeuIleAsnTrpGluValGluValAlaAsnGlnThrThrGlyGlnIleGln 253
Db      1944 AACCAAGCGCTGTGACTGGCAGGTG--GTGGCCAAAT-----GGTGAATCCAGC 1991
QY      254 IleSerValIleAspGluAspGlyValAlaIleValAlaLysAlaSerGlyValaGlnGlyThr 273
Db      1992 GTTGAACCTCGCGATGCGGATCAACAGGTGTTGCACTGCAACAGGCACTAGCGGACT 2051
QY      274 ValThrIleProSerValLysLeuTrpGlnProGlyAlaAlaTrpLeuTrpGlnLeuGln 293
Db      2052 TTGCAAGTGGTAATCCGACCTCTGGCAACCGGGTGAAAGTTACTCTATGAACGTGTGC 2111
QY      294 ValAsnIleValGlySerSerGlyAspValValAspThrTyrAsnLeuAlaThrGlyVal 313
Db      2112 GTC-----ACAGCCAAAGCCAGACAGAGTGTATATCTACCCGCTTCGCTCGGCAAT 2165
QY      314 ArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysProPheTrpPheThr 333
Db      2166 CGGTACAGTGGCAGTGAAGGCGCAACAGTTCCGTATTAACCAAAACCGTTCTTACT 2225
QY      334 GlyPheGlyLysHisGluAspThrAlaValaGlyLysGlyHisAspProAlaTrpMet 353
Db      2226 GCGTTTGGTCTCATGAATGATCGGACTTACCTGTGCAAGATTCGATACGTCGTGATG 2285
QY      354 ValHisAspPheGlnLeuMetLysTrpIleGlyValaAsnSerPheArgThrSerHisTrp 373
Db      2286 GTGCAGACCAACGATTAATGATGATGGGGCCAACTCTTACCGTACTTCGCAATTAC 2345
QY      374 ProTyrAlaGluGluValMetAspPheAlaAspArgAsnGlyIleValValIleAspGlu 393
Db      2346 CTTTACGCTGAAAGATGCTCGACTGGCGCAATGAACATGATCGTGTGATTCATGAA 2405
QY      394 ThrProAlaValGlyLeuAsnIleAlaLeu---MetGlyValSerGluSerGlyAlaPro 412
Db      2406 ACTGCTGCTGTGGCTTAACTCTCTTTAGCATTTGTTGGAAGCGGCAACAAGCGG 2465
QY      413 ---GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAlaHisIleGlnAla 431
Db      2466 AAGAACTGTACAGCAAGAGGACATCAACGGGAAACTCAGCAACCGCATTTACAGGGG 2525
QY      432 IleArgGluLeuIleAlaArgAspLysAsnHisAlaSerValValMetTrpSerIleAla 451
Db      2526 ATTAAGAGCTGATACGCGGTGCAAAACCAACCAAGCGTGTGTGTGAGTATTTGCC 2585
QY      452 AsnGlnProAlaSerHisGluAspGlyAlaArgGluTrpPheGluProLeuThrAsnLeu 471
Db      2586 AACGAACCGGATPACCGCTCGCAAGGTGACAGGAAATATTTGCGGCACCTGCGCGAGCA 2645
QY      472 ThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTrp 491
Db      2646 ACCGTAATCTGACCCGACGCGCTCGATCACCTGCGTCAATGTATATGTTTCGACGCGCT 2705

```

QY 492 GlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTyr 511  
DB 2706 CACACCGATACCTACCGCATCTCTTGATGCTGCTGCTGCAACCTTATACGATGG 2765  
QY 512 TyrSerGlnThrGlyAspLeuGluValAlaGluAlaLeuGluLysGluLeuHisGly 531  
DB 2766 TATGTCCAAAGCGCGCATTTGGAAACGCGCAGAGAGGTAAGTCTGAAAAAGAACTTCGCGC 2825  
QY 532 TrpGlnGluLysPheHisArgProIleValMetThrGluLysTyrGlyAlaAspThrLeuAla 551  
DB 2826 TGGCAGAGAAAGTGTATCGCGCATTTATCATCCGAAATACGCGGTGATACGTTAGCC 2885  
QY 552 GlyLeuHisSerIleLeuGluLysLeuProTyrSerGlnGluLysPheGluValGlnMetLeuAsp 571  
DB 2886 GGGCTGCTCATCATGATACACCGCATGTGAGTGAAGATATCATGCTGATGCGATG 2945  
QY 572 MetTyrHisArgValPheAspArgIleGluSerMetAlaGlyLysHisValTTPAspPhe 591  
DB 2946 ATGTATCACCGCGCTTTGATCGCGTACGCGCGTGTGTAACAGTATGGAATTTTC 3005  
QY 592 AlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnLysGlyValPhe 611  
DB 3006 GCGGATTTTCCGCTCCGCAAGGATATTCGCGTTCGCGTAAACAAGAAAGGATCTTC 3065  
QY 612 ThrArgAspArgLysProLysAlaAlaHisSerLeuArgAlaArgTyrThrSerIle 631  
DB 3066 ACTCGGACCGCAACCGAAGTGGCGGCTTTCTGCTGCAAAAAGCTGACTGGCATG 3125  
QY 632 Asp 632  
DB 3126 AAC 3128  
RESULT 3  
AR096536  
LOCUS AR096536 3169 bp DNA linear PAT 08-SEP-2000  
DEFINITION Sequence 5 from patent US 6008023.  
ACCESSION AR096536  
VERSION AR096536.1 GI:10025425  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 3169)  
AUTHORS Oppert M., Bouslet K. and Czech J.  
TITLE Cytoplasmic expression of antibodies, antibody fragments and  
JOURNAL Patent: US 6008023-A 5 28-DEC-1999;  
FEATURES Location/Qualifiers  
source 1..3169  
/organism="unknown"  
/mol\_type="unassigned DNA"  
ORIGIN  
Alignment Scores:  
Pred. No.: 2,2e-116 Length: 3169  
Score: 1721.50 Matches: 339  
Percent Similarity: 68.17% Conservative: 98  
Best Local Similarity: 52.89% Mismatches: 179  
Query Match: 51.33% Indels: 25  
DB: 6 Gaps: 11  
US-10-757-093-4 (1-634) x AR096536 (1-3169)  
QY 8 SerLeuLeuSerLeuAlaAlaProSerLeuGlyThrProAla-----AlaArg 23  
DB 1233 AGCGTGTGACCGTCCCTCAGACGCTTGCGGACCCAGACTTACACTGCAACGTGAAT 1292  
QY 24 HisPheProArgAsnGluMetThrGlnHisGluGlnProLeu-----Ile 38  
DB 1293 CACAAGCCGCAACACCAAGTGAACAAGAGTGAAGTGGTTCGTTGCTGCTCC 1352  
QY 39 LysValArgProGlnArgThrSerSerArgGluLeuValAsnLeuArgGlyLeuTyrLys 58

DB 1353 ATGTATCGTCTGTAGAAACCCCAACCCGTAATCAAAAACCTCGACGCGCTGGCGCA 1412  
QY PheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProThrAlaPro 75  
DB 1413 TTCAGTCTGATCCGAAATCTGTGAATTGATCAG-----CGTTGGTGGAAAGCGG 1466  
QY 76 LeuProLysGlyLeuGluCysProValProAlaSerTyrAsnAspIlePheIleSerArg 95  
DB 1467 TTACAGAAAGACCGGGCAATGCTGTGCGACGAGCTTTTAAAGATACATTCGCGCATGCA 1526  
QY 96 GluIleHisAspHisValGlyTyrValTyrTyrGlnArgGluValIleValProLysGly 115  
DB 1527 GATATTGTAATATATCGGGCAACGTCGTGTATCAGCGCGCAAGCTTTATACCGAAAGGT 1586  
QY 116 TrpSerGlnGluArgTyrLeuValArgAlaGluSerAlaThrHisGlyAspGlyIleTyr 135  
DB 1587 TGGGACGGCCAGGATTCGTGCGCTTCGATCGCGTCACTATTAACGCAAGGTGG 1646  
QY 136 ValAsnAspArgLeuValAlaGluHisValGlyIleTyrThrProPheGluAlaAspVal 155  
DB 1647 GTCAATATATCAGAAAGTGAATGAGCATCAGGCGGCTATACGCAATTTGAAGCGATGTC 1706  
QY 156 ThrGluLeuValAlaProGlyGluLysPheArgLeuThrIleGlyValAsnAsnGluLeu 175  
DB 1707 ACCCGGTATGTATTCGCGGAAAGTGAATGATACACCGTTGTGTGAACAAGAACTG 1766  
QY 176 ThrHisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysArgIle 195  
DB 1767 AACTGACAGACTATCCCGCGGAAATGTGATTAAC--GACAAACCGCAAGAAAG 1823  
QY 196 GlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaAspSerIleTyrLeuTyr 215  
DB 1824 CAGTCTTACTCATATATTTCTTAACATGCGGGAATCATGCAAGCGTAATGCTCTAC 1883  
QY 216 SerValProGlnGlnHisIleGlnAspIleThrValValThrAspVal-----AspGly 233  
DB 1884 ACCACCCGCAACCTCGGTGAGCATATCACCGTGTGACGATCGCGCAAGACTGT 1943  
QY 234 AspAsnGlyLeuLeuLeuThrGluValGluValAlaAsnGlnThrGlyGlnIleGln 253  
DB 1944 AACGACGCGCTGTATTAACGCGCAGTG--GTGGCAAT-----GGTGAATGAC 1991  
QY 254 IleSerValIleAspGluAspGlyValAlaIleValAlaLysAlaSerGlyAlaGlnGlyThr 273  
DB 1992 GTTGAACGTGATATCGCATCAACAGTGTGTCACATGCAAGCAACACTACCGGACT 2051  
QY 274 ValThrIleProSerValLysLeuTyrProGlnProGlyValAlaLysTyrLeuTyrGlnLeuGln 293  
DB 2052 TTGCAAGTGTGATACCGCACCTCTGCAACCGGGTGAAGTTATCTTATGAACGTGTC 2111  
QY 294 ValAsnIleValGlySerSerGlyAspValValAspThrTyrAsnLeuAlaThrGlyVal 313  
DB 2112 GTC-----ACAGCCAAAGCCAGACAGAGTGAATATTAACCGCTTCGCGTGGCATC 2165  
QY 314 ArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysProPheTyrPheThr 333  
DB 2166 CGGTCAAGTGTGATGAAGGCGCAACAGTCTCTATTAACCAACAAACGTTCTTACT 2225  
QY 334 GlyPheGlyLysHisGluAspThrAlaValArgGlyLysGlyLysHisAspProAlaTyrMet 353  
DB 2226 GCGTTTGTCGTGATGAAGTGGCACTTACGTCGCAAGAGATTGATACGTCGATG 2285  
QY 354 ValHisAspPheGlnLeuMetLysTyrPheIleGlyValAsnSerPheArgThrSerHisTyr 373  
DB 2286 GTGCAGACCAACCATTAATAGATGATGATGGGCAACCTCAACCTGATCTGCAATTAAC 2345  
QY 374 ProTyrAlaGluGluValMetAspPheAlaAspArgAsnGlyIleValIleAspGlu 393  
DB 2346 CTTTACGCTGAAGATGCTGCACTGGCAGAGAACATGCAATCGTGGATGATGA 2405  
QY 394 ThrProAlaValGlyLeuAsnIleAlaLeu--MetGlyValSerGluSerGlyAlaPro 412  
DB 2406 ACTGCTGCTGTGCGCTTAACCTCTCTTATGACATGTTTCGAAGCGGCAACAGCCG 2465

QY 413 ---GlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAlaHisIleGlnAla 431  
 Db 2466 AAMGAATCTTACCGCGAAGCAGTCAGCGGAACTCAAGGCACTTACAGGCG 2525  
 QY 432 IleArgGluLeuIleAlaAspAspLysAsnHisIleAspValIleMetTrpSerIleAla 451  
 Db 2526 ATTAAGAGCTGATAGCGCGTGCACAAAACCAACCGCTGGTGAATGAGATATGGCC 2585  
 QY 452 AsnGluProAlaSerHisGluAspGlyAlaArgGluIleThrPheGluProLeuThrAsnLeu 471  
 Db 2586 AACGAACCGGATACCCCTCCGCAAGTGCACGGGAATATTTCGGCCACTGGCGGAAGCA 2645  
 QY 472 ThrArgGluLeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyr 491  
 Db 2646 ACCGTAACCTCAGACCGCGCTCCGATACCTGCTCAATGATATGTTCTGGACGCT 2705  
 QY 492 GlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTyr 511  
 Db 2706 CACACCGATACCATCAGCGATCTCTTGATGTGTGGTGGCTGAACGCTTATACGGATGG 2765  
 QY 512 TyrSerGlnThrGlyAspLeuGluGluAlaGluAlaIleuGluLysGluLeuHisGly 531  
 Db 2766 TATGTCGAAAGCGCGCATTTTGAAACGCGAGAGAGGTTACTGAAAAGAACTTCTGGCC 2825  
 QY 532 TrpGlnGluLysPheHisArgProIleValIleMetThrGluIleValIleAspThrLeuAla 551  
 Db 2826 TGGCAGAGAAATCGATCATCGCGATATCATCAGCAATACCGCGGTGATACGTTAGCC 2885  
 QY 552 GlyLeuHisSerIleLeuGlyLeuProTrpSerGluGluIleValGlnMetLeuAsp 571  
 Db 2886 GGGCTGACTCAATCATACCGACCATGTGGAGTGAAGATATCATGTGCATGGCTGGAT 2945  
 QY 572 MetTyrHisArgValPheAspArgIleGluSerMetAlaGlyGluHisValIleTrpAsnPro 591  
 Db 2946 ATGTATCAGCCGCTTTTGAATCGGTGAGCGCGTGTGCTGTAACAGATGAAATTTTC 3005  
 QY 592 AlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsnLysGlyValAlaPhe 611  
 Db 3006 GCCGATTTTGCACCTCGCAAGGCAATTTGCGCGTTGCGGTGAACAAGAAAGGATCTTC 3065  
 QY 612 ThrArgAspArgLysProLysAlaIleAlaHisSerLeuArgAlaArgTrpThrSerIle 631  
 Db 3066 ACTCGGACCGCAACCGAGTGGCGGCTTTTCTGTGCAAAAACGCTGACTGCGCANG 3125  
 QY 632 Asp 632  
 Db 3126 AAC 3128  
 Db 3126 AAC 3128

RESULT 4  
 AR373248 3169 bp mRNA linear PAT 18-DEC-2003  
 LOCUS AR373248  
 DEFINITION Sequence 5 from patent US 6602688.  
 ACCESSION AR373248  
 VERSION AR373248.1 GI:40075256  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 Unclassified.

REFERENCE  
 1 (bases 1 to 3169)  
 Oper, M., Bosslet, K. and Czech, J.  
 Cytoplasmic expression of antibodies, antibody fragments and  
 antibody fragment fusion proteins in E. coli  
 Patent: US 6602688-A 5 05-AUG-2003;  
 Location/Qualifiers  
 1. 3169  
 /organism="unknown"  
 /mol\_type="mRNA"

ORIGIN  
 Alignment Scores: 2.2e-116 Length: 3169  
 Pred. No.: 1721.50 Matches: 339  
 Score:

Percent Similarity: 68.17% Conservative: 98  
 Best Local Similarity: 52.89% Mismatches: 179  
 Query Match: 51.33% Indels: 25  
 DB: 6 Gaps: 11

US-10-757-093-4 (1-634) x AR373248 (1-3169)

QY 8 SerLeuLeuSerLeuAlaIleProSerLeuGlyThrProAla-----AlaArg 23  
 Db 1233 AGCGTGTAACCGTGCCTCCAGACACTTGGGACCCAGACCTACACTGCAACGTGAAT 1292  
 QY 24 HisPheProArgAsnGluMetThrGlnHisGluGlnProLeu-----Ile 38  
 Db 1293 CACAAACCCAGCAACCAACGAGTGGACAAAGAGTGAAGTGGCTTGTGTTCTGCTCC 1352  
 QY 39 LysValArgProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpLys 58  
 Db 1353 ATGGTACGTCCTGTAGAAACCCCAACCGCTGAATCAAAAACTGACGCGCTGTGGCA 1412  
 QY 59 PheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProThrAlaPro 75  
 Db 1413 TTCAGTCTGATCGCAAAACTGTGAATGTATCAG-----CGTTGTGGAAACGCG 1466  
 QY 76 LeuProLysGlyLeuGluLysProValProAlaSerTyrAsnAspIlePheIleSerArg 95  
 Db 1467 TTACAAAGAAAGCGCGCATTCGCTGTGCGACGACAGTTTAAAGTACGTTCCCGCATGA 1526  
 QY 96 GluIleHisAspHisValGlyTrpValTyrTyrGlnArgGluValIleValProLysGly 115  
 Db 1527 GATATTCTGAATTAATCCGGCAACGCTGTGTACGCGCGGAAGTCTTATACCGAAAGGT 1586  
 QY 116 TrpSerGlnGluArgTyrLeuValArgAlaGluSerAlaThrHisIleGlyArgIleTyr 135  
 Db 1587 TGGGACGCGCAAGCTATCTGCTGCTTTCATGCGGTCACTCATTCACGCAAGTGTGG 1646  
 QY 136 ValAsnAspArgLeuValAlaGlnHisValGlyGlyTyrThrProPheGluAlaAspVal 155  
 Db 1647 GTCATATATCAGAAAGATGATGAGCATCAGGCGGCTATACCCATTGGAACCGCATGTC 1706  
 QY 156 ThrGluLeuValAlaProGlyLysPheArgLeuThrIleGlyValAsnAsnGluLeu 175  
 Db 1707 AGCGCGTATGTAATGCGCGGAAAGTGCATATCACCGTTGCTGTGAACAACAACGTC 1766  
 QY 176 ThrHisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysArgIle 195  
 Db 1767 AACGTCAGACATATCCCGCGGGAATGATTAAC--GACGAAGCGCAAGAAAAAG 1823  
 QY 196 GlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAlaArgSerIleTrpLeuTyr 215  
 Db 1824 CAGTCTTACTTCATATATTTCTTAATACTATGCGGGATCCATCGACGTAATGCTTAC 1883  
 QY 216 SerValProGlnGlnHisIleGlnAspIleThrValAlaThrAspVal-----AspGly 233  
 Db 1884 ACCACGCGCAACACTGGGTGACAGATATCACCGTGTGACGATCGCGCAACACTGT 1943  
 QY 234 AspAsnGlyLeuIleAsnTyrGluValGluValAlaAsnGlnThrThrGlyGlnIleGln 253  
 Db 1944 AACCAAGCGTGTGATCGCGAGTG--GTGGCAAT-----GATGATATCGAC 1991  
 QY 254 IleSerValIleAspGluAspGlyAlaIleValAlaLysAlaSerGlyAlaGlnGlyThr 273  
 Db 1992 GTTGAACGCGGATGCGGATCAACAGGTGGTTCGCAATGTCGCAAGGACTAGCGGACT 2051  
 QY 274 ValThrIleProSerValLysLeuTrpGlnProGlyAlaAlaTyrLeuTyrGlnLeuGln 293  
 Db 2052 TTGCAAGTGGTAATCCGACCTTGGCAACCGGTGAAGTTATCTTATGAACCTGTGC 2111  
 QY 294 ValAsnIleValGlySerSerGlyAspValValAspThrTyrAsnLeuAlaThrGlyVal 313  
 Db 2112 GTC-----ACAGCAAAAGCGACAGAGTGTATATCTACCCGCTTCCGCTCGCATTC 2165  
 QY 314 ArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLysPheProPheThrThr 333



/note="modified form mut3 from Aequorea victoria GFP found  
in GenBank Accession number M62653"  
/codon\_start=1  
/transl\_table=11  
/product="green fluorescent protein"  
/protein\_id="A03194.1"  
/db\_xref="GI:30523397"  
/translation="MSKGEELFTGVVPIVELDGDVNGHKFSVSGEGDPAIVGKLT  
KFTCTGKLPVPPPTVTPGCVGCCAARYPDHMKDHPKSMRPGYUQERTIPKQ  
DKNYKRAEYFERGDTLVNRIELKQIDFKEDGNILGHKLEFYNYSNRYIIMADKQKQ  
IKNFKIRHNIEDGSVOLADHYQONTPIGDVPLLPDNLHLSQSALSKDPNEKRDM  
VLEFVTAAGITGHMDELYK"  
/complement(7966. .8826)  
/gene="bla"  
/complement(7966. .8826)  
/gene="bla"  
/codon\_start=1  
/transl\_table=11  
/product="beta-Lactamase"  
/protein\_id="A03195.1"  
/db\_xref="GI:30523398"  
/translation="MSIQHRYVALIPPAFCIPVAHPETLVKKAEDQLGARVY  
IETDLSGKILESFRPEERPMWTEKVLICGAVSRIDAGQQLGRIRHYSQNDLVE  
YSPVTEKHLTDGMTVELCSAATMSDNTAAILTTIGPKELTAFLHNMGDVTRL  
DRWEPLNEAIIPMDERDTTPVAMATLRLKLTGLIELTLASPKQLIDMEADKVAQPL  
LRGALPAGWFIADKSGAGERGSRGIIAALGPDPKPSRIIVITVTSQATMDERNQIA  
ELGASLIKHM"

## ORIGIN

## Alignment Scores:

Pred. No.: 1,42e-115 Length: 9037  
Score: 1718.50 Matches: 340  
Percent Similarity: 69.00% Conservative: 94  
Best Local Similarity: 54.05% Mismatches: 172  
Query Match: 51.24% Indels: 23  
DB: 12 Gaps: 10

US-10-757-093-4 (1-634) x AY286001 (1-9037)

QY 18 G1YThProAlaIaIaRHisPheProArGAsnGlu--MetThRgInHisGluInProL 37  
Db 6597 GGAATTCCTGGACCCCGGGGATCCACAGAAACGCTATGACCATGATTACGCCAAGCT 6538  
QY 37 eu-----1lelyVaIaArPProGlnArGThrSerSerArGluLeuVaIa 52  
Db 6537 TGGCTCAGAGTACGTCCCTTATGCTCGCTCTAGAAACCCCAACCCGGAATCAAAA 6478  
QY 52 snleuAspGlyLeuTTPlysPheAlaLeu-----AlaSerGlyLeuAsnAspThra 69  
Db 6477 AACTTCAGCGCCCTGTGGCATTCAGTCTGGATCGCAAAAACCTGGAAATGATCAG---- 6422  
QY 69 lagInProTThrAlaProLeuProlysglyLeuGluCyProValProAlaSerTyra 89  
Db 6421 --CGTTGGTGGAAAGCCGCTTCAAGAAAGCCGGCAATTCCTGTGGCCAGGAGTTTAA 6364  
QY 89 snAspIlePheIleSerArGluIleHisAspHisValGlyTrpValIyTyrgInArG 109  
Db 6363 ACATCAGATTCCGCGCAATTCGTAATATAGCGGCAACGCTCGGTATCAACGGG 6304  
QY 109 lVaIleValaPProlysglyTPSerGlnGluArGlyLeuValaArGAlaGluSerAlaT 129  
Db 6303 AAGTCTTTATACGGAAGGTGGGAGGCCAGCGTATCGTGCCTTTTCGATCGCGTCA 6244  
QY 129 hThiHisGlyArGlyIeTyraIaAsnAsnArGLeuValaIaGluHisValaGlyTyrt 149  
Db 6243 CTCATTACCGCAAAAGTGGGTCAATTAATCAGAAAGATGAGCATTCAGGGGGCTATA 6184  
QY 149 hPProheGluIaAspValThrGluLeuValaIaProGlyGluLysPheArGLeuThr 169  
Db 6183 CGGCATTTAAGCGCATGTCAGCCGTAATGTTATTCGCGGAAAAAGTGAACGATCAACG 6124  
QY 169 lGlyValaAsnGluLeuThiHisGluThrIleProProGlyLysIleThrThiGlyA 189  
Db 189 :::

Db 6123 TTTGTGNGAACAAACGAACCTGAACCTGACAGACTATCCCGCGGGAATGTGATTACC---G 6067  
QY 189 snAlaThrGlyValaGllleGlnThrTyrgInHisAspPheTyraNtyrAlaGlyLeuA 209  
Db 6066 ACGAAACCGCAAGAAAGCAAGCTCTTACTTCCATATTTCTTAACCTATGCGGAAATCC 6007  
QY 209 laArGserIleTTPleuTyrsSerValProGlnGlnHisIleGlnAspIleThrValaIaT 229  
Db 6006 ATGCGAGCTAATGCTCTTACACCAACCCCAACACCTGGGTGACGATATCACCGTGTGA 5447  
QY 229 hTrAspVal-----AspGlyAspAsnGlyLeuIleAsnTyrgluValaIaAsnG 247  
Db 5946 CGCATGTCGCGCAAGCTGAACCAACGCGCTGTGTACTGACGAGGTG---GTGGCCAAAT- 5891  
QY 247 lnrThrThrglyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAlaLysA 267  
Db 5890 -----GCTGATGTCACCGTGAATCTCGTATGCGGATCAACAGGTGTGCAACTG 5839  
QY 267 laSerGlyAlaGlnGlyThrValThrIleProSerValLysLeuTPGInProGlyAlaA 287  
Db 5838 GACAAAGCACTAGCGGGAATTTGCAAGTGTGATCCGACCTTGGCAACGGGTGAAG 5779  
QY 287 laTyLeuTyrgluLeuGlnValaAsnIleValGlySerSerGlyAspValaIaAspThrT 307  
Db 5778 GTTATCTTATGAACCTGTGCCTC-----ACAGCCAAAGCCAGACAGATGTGATATCT 5725  
QY 307 yTrAsnleuAlaThrglyValaArGThrValaLysValaIaGlySerGlnPheleuIleAsnG 327  
Db 5724 ACCGCTTCGCTCGGCACTCGGTCAGTGCAGTGAAGGCGCAACAGTTCGATTAAAC 5665  
QY 327 lYAspProheTyrrPheThrglyPheGlyLysHisGluAspThrAlaValaArGlyLysG 347  
Db 5664 ACAAACCGTTACTTACTTATGCTGCTTTGGTCGTCATGAATACCGGACCTTACGTGCAAG 5605  
QY 347 lYHisAspProAlaTyrmecValHisAspPheGlnLeuMetLysTrpIleGlyAlaAsn 367  
Db 5604 GATTCCATTAACGTGCTGATGTGTCAGCACCGACCATTAATGACTGTGATGGGCCCAACT 5545  
QY 367 erPheArGThrSerHisTyrrProTyrrAlaGluGluValaMetAspPheAlaAspArG 387  
Db 5544 CTTACCGTACTCGCACTTACCTTACCTTACGTAAGAGATGCTCGACGTGGGACATGAACATG 5485  
QY 387 lYIleValaIleAspGluThrProAlaValaGlyLeuAsnIleAlaLeu---MetGlyV 406  
Db 5484 GCATCGTGTGATGATGAATGAATGCTGCTGCTGCTTAACTCTTTAAGCATTTGCT 5425  
QY 406 alSerGluSerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLysThrG 425  
Db 5424 TCGAAGCGGCAACAAACCGAAGAACTGTACAGCAAGAGCAGTCAACCGGGAAACTC 5365  
QY 425 lngluAlaHisLysGlnAlaIleArGluLeuIleAlaArGAspLysAsnHisAlaSerV 445  
Db 5364 AGCAAGCGCACTTACAGCGGATTAAGACTTAAGCGGTGAACAAACCCACCAAGCG 5305  
QY 445 alValMetTPserIleAlaAsnGluProAlaSerHisGluAspGlyAlaArGlyTyrrP 465  
Db 5304 TGGTATGTGAGTATTCGCAACGAACCGGATACCGCTCGCAAGGTGACGGGAATAT 5245  
QY 465 heGluProLeuThrAsnleuThrArGlnLeuAspProThrArGProIleThrPheAlaA 485  
Db 5244 TCGCGCACCTGCGGAGCAAGCAACGCGTAACTGACCCGCGGTCCGATCACCTCGTCA 5185  
QY 485 snValGlyThralaThrTyrgluLeuAspArgIleSerAspLeuPheAspValaIserCyA 505  
Db 5184 ATGTATATGTTTCGCAAGCTCACACCGATACCATCAGCATTTCTTATGTGCTGTGCC 5125  
QY 505 lAsnArGtyrPheGlyTyrrTyrsGlnThrGlyAspLeuGluIaGluAlaIaAlaL 525  
Db 5124 TGAACCGTTATTAACGATGTATGTCCAAAGCGGAGATTGGAACCGGACGAAGAGTAC 5065  
QY 525 euGluLysGluLeuHisGlyTyrgInGluLysPheHisArGProIleValaMetThrGluT 545  
Db 5064 TGGAAAAAGAACTTCTGGCTTGGCAGAGAACTCATCAGCGGATTCATCACCAGAT 5005

QY 545 YRG1YALAAPHThleuAlaGlyLeuHisSerIleLeuProTIPSerGluGlp 565  
 DB 5004 AGGCGCTGATACGTTAGCCGCGCTCACTCAATGACACCGACATGCGAGTGAAGAGT 4945  
 QY 565 heGlnValAlnMetLeuAspMetTyrHisArgValPheAspArgIleGluSerMetAlaG 585  
 DB 4944 ATCAGTGTGCATGCGCGATATGTATCACCAGCGCTTTGATGCGTCAAGCGCGTGTG 4885  
 QY 585 YGluHisValTTPaaphPheAlaAspPheGlnThrAsnLeuGlyTlleIleArgValAspG 605  
 DB 4884 GTGACAGGATATGAAATTCGCGCATTTTCGCGACCTCGCAAGGCATATTCGCGTGGCG 4825  
 QY 605 YAsnLeuValGlyValPheThrArgAspArgGlyProLysAlaAlaHisSerLeuA 625  
 DB 4824 GTTACAGAAAGGAGCTTCATCGCGACCGCAACCGAAGTCGCGCTTTCTGTGTC 4765  
 QY 625 rGAlaArgTTPThrSerIleAsp 632  
 DB 4764 AAAAAGCTGACTGCGCATGAC 4742

RESULT 6  
 AY237647/c 10313 bp DNA circular SYN 22-MAR-2003  
 LOCUS Cloning vector pHRGFPUS, complete sequence.  
 DEFINITION AY237647  
 ACCESSION AY237647.1 GI:29150587  
 VERSION  
 KEYWORDS Cloning Vector pHRGFPUS  
 SOURCE Cloning Vector pHRGFPUS  
 ORGANISM other sequences; artificial sequences; vectors.

REFERENCE  
 AUTHORS 1 (sites)  
 TITLE Quandt, J. and Hynes, M.F.  
 JOURNAL Versatile suicide vectors which allow direct selection for gene  
 MEDLINE replacement in gram-negative bacteria  
 PUBMED Gene 127 (1), 15-21 (1993)  
 93252274  
 8486283

REFERENCE  
 AUTHORS 2 (sites)  
 TITLE Metcalf, W.W. and Wanner, B.L.  
 JOURNAL Construction of new beta-glucuronidase cassettes for making  
 MEDLINE transcripional fusions and their use with new methods for allele  
 PUBMED replacement  
 9328120  
 9328120 (1), 17-25 (1993)

REFERENCE  
 AUTHORS 3 (bases 1 to 2642; 6610 to 7326)  
 TITLE Ouahran, Bettache, S., Porte, F., Teyssier, J., Liautard, J.P. and  
 Kohler, S.  
 JOURNAL PBRI-GFP: a broad-host-range vector for prokaryotic promoter  
 MEDLINE studies  
 PUBMED Biotechniques 26 (4), 620-622 (1999)  
 99275364  
 10343896

REFERENCE  
 AUTHORS 4 (bases 1 to 10313)  
 TITLE Soares-Ramos, J.R.L., Hungria, M. and Pedrosa, F.O.  
 JOURNAL Monitoring Azospirillum-bacterial interactions using the gfp and gusa  
 MEDLINE genes constitutively expressed from a new broad-host range vector  
 PUBMED J. Biotechnol. 97 (3), 243-252 (2002)  
 22079789  
 12084480

REFERENCE  
 AUTHORS 5 (bases 1 to 10313)  
 TITLE Soares-Ramos, J.R.L., Souza, E.M. and Pedrosa, F.O.  
 JOURNAL Submitted (14-FEB-2003) Department of Biochemistry and Molecular  
 MEDLINE Direct Submission  
 PUBMED Submitted (14-FEB-2003) Department of Biochemistry and Molecular  
 12084480

COMMENT  
 Biology, Universidade Federal do Parana - UPR, Centro Politecnico,  
 Curitiba, PR 81531-990, Brazil  
 Mobilization (mob) and replication (rep) regions are derived from  
 the broad-host-range vector PBRI isolated from Bordetella  
 bronchiseptica. The expression of uidA and gfp genes is under  
 control of gentamycin resistance aacCI promoter. The pgen promoter  
 and part of the gentamycin gene (aacCI) were isolated from pJQ200KS

FEATURES as a BglII/BamHI fragment.  
 source Location/Qualifiers  
 1..10313  
 /organism="Cloning Vector pHRGFPUS"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:22838"  
 /note="Broad-host-range vector expressing GFP"  
 1..2642  
 /note="mob and rep regions from broad-host-range vector  
 PBRI-GFP as deposited in GenBank Accession number  
 AF110459"  
 complement(2982..3376)  
 /gene="aph"  
 /note="NPTII; derived from vector pMW6"  
 complement(2982..3376)  
 /gene="aph"  
 /codon\_start=1  
 /transl\_table=1  
 /product="aminoglycoside-3'-O-phosphotransferase"  
 /protein\_id="AA063174.1"  
 /db\_xref="GI:29150589"  
 /translation="MIBDGLHAGSPAMVERLFGYDAAQOTIGSDAAYFLSLAQR  
 PVLFWKTLDSALNELQDEARLSWATGVCALVDVITAGRWMLGSRPGDGL  
 LSSHLAPAEKVSIMADAMKRLHTLPATCPDPAQKHRLERARTMEAGLVQDGLDE  
 EHOGAPAELEFARLAKAMPDGEDLVYTHGDACTPIIMVNGRFSGFIDCGRLGVADRY  
 ODIALATRDIAELGEMVADRFLVYGTAAAPDSQSLAFYRLDDEF"  
 complement(3865..3914)  
 /function="neomycin and kanamycin resistance aph (NPTII)  
 gene promoter region"  
 complement(4705..6517)  
 /gene="uidA"  
 /note="promoterless uidA gene from pMW6"  
 complement(4706..6517)  
 /gene="uidA"  
 /codon\_start=1  
 /transl\_table=1  
 /product="beta-glucuronidase"  
 /protein\_id="AA063175.1"  
 /db\_xref="GI:29150589"  
 /translation="MWRPEYTPTRREIKLDDGMAFSLDRENGCIDORMMWESALQESRA  
 IAVPSFNDQFADADIRNAGVNWQREVFITKAGAGQIVLPDAVTHYKGVWNO  
 EWEHOGGTPEADAVTPVILAGKSVIRIVCNVNEIMQTIIPGAVITDENKQSY  
 FHDFFNYAGIHSVMLYTPNTWVDITVTHVADCNHASVDMQVANGDVSVELRD  
 ADOOVAVATGCGTSGTLQVNPVHLMQBGELVELCVTASQTECDIPLRVGIRSVAY  
 KEQRIINRKEPPTFGFGRHEDADIRKGFNDVILVHDLMDWIGANSVRSHPYA  
 EEMLWADHEGIVVIDETAVGPNISLIGFPAKPKKLYSEBAVNGTQAHAIQAI  
 KELIARDKNHPSVVMSSINERDTRPQGRAYFAPLAEXTRLDPTRLPTCVNWFCD  
 AHTDIISDLFDVLCINRYGYVVGSDLETAKVLEKELIAOEKIHPIITTEYGVN  
 TLAGHSWYTDWMSSEYOCAMDYHRVDRSAVVEQVWNFADPATSQGILRVGYN  
 KKGITTRDKPKSAFLLOKRWTKMNFGRKPPQGGKQ"  
 complement(6610..7326)  
 /gene="gfp"  
 /note="promoterless gfp gene from PBRI-GFP; modified form  
 mut3 from Aequorea victoria GFP as deposited in GenBank  
 Accession number M62653"  
 complement(6610..7326)  
 /gene="gfp"  
 /note="modified form mut3 from Aequorea victoria GFP as  
 deposited in GenBank Accession number M62653"  
 /codon\_start=1  
 /transl\_table=1  
 /product="green fluorescent protein"  
 /protein\_id="AA063176.1"  
 /db\_xref="GI:29150590"  
 /translation="MSKGELEFTGVVPLVLELDGVNKHKFSVSGEGEDATYKLTLL  
 KILCTTGKLPVMPPLVTTFYGVQCFARYPDHMKQHDFFKAMPEGYQESTTIPKQD  
 DQNYTKRAVKEGDTLVNRIELKGIQDFEDSNILGKLELVNYSNIVYIMDKQNG  
 IKVNEKIRHNIEDGVSGLADHYQNTPIEDGVLVLPDNYLSTQSLSDPBEKRDHM  
 VLEFVTAAGITGHGMDLYK"  
 7363..7860  
 /note="from pJQ200KS"  
 complement(7865..7893)  
 /note="-35 and -10 promoter regions of the gentamycin

misc\_feature  
 promoter



```

misc_feature      resistance aacC1 gene from pJ0200KS"
                   /note="from pJ0200KS"
                   complement(9242..10102)
gene              /gene="bla"
                   /note="derived from pBBR1-GFP"
                   complement(9242..10102)
CDS               /gene="bla"
                   /codon_start=1
                   /transl_table=11
                   /product="beta-lactamase"
                   /protein_id="AA063177.1"
                   /db_xref="GI:29150591"
                   /translation="MSIQHFRVALIPEFAFCLPVAPHEITLVKVAEDQLGARVGY
                   IELDLSGKILSFEPREPRPMWSTFKVLGAVLRIDAGQSLGRIRIHSQNDLVE
                   YSPVTEKHITDGMETVRELCSAATMSDNTKANLLTTIGREPTELTAFLNMGNVRL
                   DREPELNEAIIPDERDITTPVAMATTKRLTGLIELTLASROQLIDMEADIVAGPL
                   LRSALPAGWPIADKSGAGERSKGIIALGPDGKPSRIIVITYTGSQATMDERNQIA
                   EIGASLIDKIM"

ORIGIN

Alignment Scores:
Pred. No.:      1.69e-115      Length:      10313
Score:          1718.50        Matches:      340
Percent Similarity: 69.00%      Conservative: 94
Best Local Similarity: 54.05%      Mismatches: 172
Query Match:    51.24%         Indels:       23
DB:             12             Gaps:        10

US-10-757-093-4 (1-634) x AY237647 (1-10313)

QY 18 GLYThrProAlaIalaRghisPheProArgAsnGlu--MetThrGlnHisGluGlnProL 37
Db 6597 GGAATTCCTGCAAGCCCGGGGATCCACAGAAACAGCTAGACCATGATTACCCCAAGCT 6538
QY 37 eu-----11eLysValAlaArgProGlnArgThrSerSerArgGluLeuVala 52
Db 6537 TGGCTGACAGTCAGTCCTTATGATGTCGTCCTGTAGAAACCCCAACCCGTGAATCAAAA 6478
QY 52 snLeuArgIleuThrLysPheAlaLeu-----AlaSerGlyLeuAsnAspThra 69
Db 6477 AATTCACAGCGCTGTGGGCACTTCAGTCCGAAACCTGTGAATTTGATGACAG-- 6422
QY 69 IagInProTrpThrAlaProLeuProLysGlyLeuGlyCysProAlaSerTyrA 89
Db 6421 --CGTTGGTGGAAAGCGCTTCAAGAAAGCCGGCAATGGCTGTGCCAGCACTTTTA 6364
QY 89 snAspIlePheIleSerArgGluIleHisAspHisValGlyTyrValTyrTyrGlnArg 109
Db 6363 ACATGACATTCGCCGAGTGCAGATATTCGTAATATGCGGGCAACGTCGTGATCACGCG 6304
QY 109 luValIleValProLysGlyTyrSerGlnGluArgTyrLeuValaGluUserAlaT 129
Db 6303 AAGTCTTTATACCGGAAGGTTGGGCAAGCGCATTCGTCGCGCTTTCGATGCGCTCA 6244
QY 129 hrHisHisGlyArgIleTyrValaAsnAspArgLeuValaIaGlnHisValGlyTyrT 149
Db 6243 CTCATTACCGCAAAAGTGGGTCAATATTCAGAAATGATGACATCAGCGCGGCTATA 6184
QY 149 hrProPheGluAlaAspValThrGlyLeuValaAlaProGlyLysPheArgLeuThrI 169
Db 6183 CGGCATTTCAGACGATGTCACCCCGTATGTTATTCGCCGGAAMAATGTATCATCACCG 6124
QY 169 lglValaAsnAspGlyLeuThrHisGlyThrIleProProGlyLysIleThrThrGlyA 189
Db 6123 TTGTGTGAACAAGCTGAACCTGACGATATCCCGCGGGAATGAGTATTAAC--G 6067
QY 189 snAlaThrGlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuA 209
Db 6066 ACGAANAACGGCAAGAAAGCACTCTTATCTTCATGATTTCTTTAATCTAGCCGGAATCC 6007
QY 209 lArgSerIleTrpLeuTyrSerValProGlnGlnHisIleGlnAspIleThrValaValT 229
Db 209 lArgSerIleTrpLeuTyrSerValProGlnGlnHisIleGlnAspIleThrValaValT 229

```

```

Db 6006 ATCGCAGCGTAATGCTCTACACCAACCGCGAACAAGCTGGTGGACGATATCACCGGTGTGA 5947
QY 229 hrAspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGluValaGluValaAsnG 247
Db 5946 CGCATGTCCGCGAAGCTGTATACCAACGCGCTCTGTGTACTGGCAGGCTG--GTGGCCAAAT- 5891
QY 247 lnrThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValaAlaLysA 267
Db 5890 -----GATATGTACAGCGTGAAGCTCGATATCGGATCAACAGCGTGTTCACACG 5839
QY 267 lAspGlyAlaGlnGlyThrValThrIleProSerValLysLeuTrpGlnProGlyAlaA 287
Db 5838 GACAAAGCACTAGCGGGAAGCTTTCAGATGATCCGACCTTCGGCAACCGGCGTGAAG 5779
QY 287 lATyrLeuTyrGlnLeuGlnValaAsnIleValaGlySerSerGlyAspValValAspThrT 307
Db 5778 GTTATCTTATGAACCTGTGCGTC-----ACAGCCAAAGCCAGACAGATGTGATATCT 5725
QY 307 yrAsnLeuAlaThrGlyValaArgThrValaValaValaGlySerGlnPheLeuIleAsnG 327
Db 5724 ACCGCTTCGCGTCGCGCATCCGGTCAGTGCAGTGAAGGCCAAGCTTCCTGATTAAC 5665
QY 327 lYlySerProPheTyrPheThrGlyPheGlyLysHisGlyAspThrAlaValaArgGlyLysG 347
Db 5664 ACAAAACCGTTCTTACTTACTGCTTGGTGTGTCATGAAATCGGACCTTACGTGGCAAG 5605
QY 347 lYHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTrpIleGlyAlaAsnS 367
Db 5604 GATTGATTAACGTCCTGATGTGTGACGACACCGCATTAATGACCTGATTTGGGCGCAACT 5545
QY 367 xPheArgThrSerHisTyrProTyrAlaGluGluValaMetAspPheAlaAspArgAsnG 387
Db 5544 CCAACGCTACCTCGCATTCCTTACCTTACGTTAGAGAGATGCTGACGTGGCGCAATGAACATG 5485
QY 387 lYlIleValIleLeuArgIleThrProAlaValaGlyLeuAsnIleAlaLeu---MetGlyV 406
Db 5484 GCATCGTGTGATGATGTAAGAACTGCTGTGCTGCTTAACTCTCTTAAAGCAATGGT 5425
QY 406 alSerGluSerGlyAlaPro--GlnThrPheThrProAspAlaIleAsnAspLysThrG 425
Db 5424 TCGAAGCGGCAACAAGCGAAAGAACTGTACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 5365
QY 425 lngIluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAlaSerV 445
Db 5364 ACGAAGCGCATTAACGCGGATTAAGAGCTATAGCGGTGACAAACCAACCAACCAAGCG 5305
QY 445 alValMetTrpSerIleAlaAsnGluProAlaSerHisGlyAspGlyAlaArgGlyTyrP 465
Db 5304 TGGTATGTGGAGTATTTGCCAAGCAACCGGATACCCGTCGCAAGGTGCACGGGAATAT 5245
QY 465 heGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaA 485
Db 5244 TCGCGCCACTGGCGGAAGCAAGCGGTAATGAGCCGCGGCTCGATCACCTGCGCTTA 5185
QY 485 snValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAspValaSerCysI 505
Db 5184 ATGTATATGTTTCGCAAGCTCACACCGAATACATACGATCTCTTATGTGTCGTGCC 5125
QY 505 lAsnArgTyrPheGlyTyrTyrSerGlnThrGlyAspLeuGlnGluAlaGluAlaAlaL 525
Db 5124 TGAACCGTTATACCGATGTGTATGTCCAAAGCGCGGATTTGAAACCGCGAAGAGGTAC 5065
QY 525 euGluLysGlyLeuGlnHisGlyTyrGlnGlyLysPheHisAspProIleValaMetCtnGluT 545
Db 5064 TGAATAAACAATCTTCGCGCTGACAGAGAACTGATCAAGCGGATTTATCATCACCGAAT 5005
QY 545 yrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGlnGluP 565
Db 5004 ACGGCGGTGATACGTTAGCGGCGCTGCACTCAATGTACACCGCAATGTGAGTGAAGAAT 4945
QY 565 heGlnValaGlnMetLeuAspMetTyrHisArgValaPheAspArgIleGluSerMetAlaG 585
Db 4944 ATCAGTGTGCATGCTGTGATATGATCAACCGGCTCTTGTATGCGGTGACGCGCGTCTGCG 4885

```



QY 585 |YGLIHuIeVal|TTPanpheAlaaspPheGlnThrAnLeuGlyIleIleArgValAspG 605  
 DB 4884 GTGAACAGGATGGAAATTCGCCGATTTCCGACCTCGCAGACCATATTCGCCGCTTGCCG 4825  
 QY 605 |YAsnLyLeYsGlyVal|PheThrArgAspArgLySProLyAlaAlaAlaHisSerLeuA 625  
 DB 4824 GTTACAGAGAAAGGATCTTTCACCTCGGACCGCAACCGAAGTGTGGCGCTTCTCTGTC 4765  
 QY 625 rGAlaArgTTrpThrSerIleAsp 632  
 DB 4764 AAAACGCTGACTGCATCAAC 4742

RESULT 7  
 AY237648/c 11973 bp DNA circular SYN 19-APR-2003  
 LOCUS AY237648 11973 bp DNA circular SYN 19-APR-2003  
 DEFINITION Cloning vector pHR50, complete sequence.  
 ACCESSION AY237648  
 VERSION AY237648.1 GI:30025555  
 KEYWORDS  
 SOURCE Cloning vector pHR50  
 ORGANISM Cloning vector pHR50  
 other sequences; artificial sequences; vectors.  
 REFERENCE 1 (bases 1 to 11973)  
 AUTHORS Ramos,H.J.O., Soares-Ramos,J.R.L., Souza,E.M. and Pedrosa,F.O.  
 TITLE Submitted (14-FEB-2003) Department of Biochemistry and Molecular  
 BIOLOGY, Universidade Federal do Parana - UFPR, Centro Politecnico,  
 Curitiba, Parana 81531-990, Brazil  
 FEATURES  
 source  
 1. 11973  
 /organism="Cloning vector pHR50"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:227048"  
 1. 2642  
 /note="mob and rep regions from broad-host-range vector  
 pBBR1-GFP deposited in GeneBank Accession number AF110459"  
 complement(2982..3776)  
 /gene="aph"  
 /note="synonym: nptII; derived from pMM6"  
 complement(2982..3776)  
 /gene="aph"  
 /note="NPTII kanamycin resistance gene"  
 /codon\_start=1  
 /transl\_table=11  
 /product="aminoglycoside-3'-O-phosphotransferase"  
 /protein\_id="AAP04589.1"  
 /db\_xref="GI:30025556"  
 /translation="MIEODGLHAGSPAAWVERLFGYDMAQOTTIGSDAAVFRLSAQR  
 PULFVKTDSGALNELQDEARLSMLATTCPCPAALVDVVTAGRMDLLIGVPODL  
 LSHLLAPAEKVSIMADMRRLHTLDPATCFPDQAKRIRERATREAGIYDODDDE  
 EHOGCLPAELFARKAMPDEGLLVTHGACLPNIMVENGSRPSGFIIDCGRLGVAURY  
 ODIALTRDIAEELGEMADRFVLVGIAPDSQRIAFRLDEFF"  
 complement(4706..6517)  
 /gene="uidA"  
 /note="derived from pMM6; expressed from gentamycin  
 resistance aacCI promoter"  
 complement(4706..6517)  
 /gene="uidA"  
 /codon\_start=1  
 /transl\_table=11  
 /product="beta-glucuronidase"  
 /protein\_id="AAP04590.1"  
 /db\_xref="GI:30025557"  
 /translation="MYRPVETTPREIKLGLMAFSLDRENCIDIDRMESALQESRA  
 IAVGSPNDQFADADIRYAGNAYVOREVFIPIGMAGORIVLFDVATHYGKWNQ  
 EWNEHOGGYTPFADAVTPYIAGSVITICVNNELMOTIPGMVITDENGKKOSY  
 FHDFFNYAGIHRSVMLYTTPTWVDITVTHAOCNASHVDVMOVANGDYSVELRD  
 ADOQVATGCGTSGTLQVNNPHMGOEGEGLVELCTTAAGTCDCDIYPLRVGRSTAV  
 KGGQFLINRCPYFTFGGRHEDADLKGKGDVNLVMDHMLDMIGANSRISHIYA  
 EEWLWADHEGIVIDETAVALNLSLIGFGEFANVAPKEXLEYSEAVANGSTIQOHLAI  
 KELIARDKHPSPVVMVSIANEPTTRQGAERYAPLAELTRKLDPTPRITCVNMVCD  
 AHTDTISDLFDVLCLNRYGMYVQSGDLETAKEVLEKELLAMQEKLHQPITITREYVD

gene  
 TAGLHSMYTDMMSEBYOCAMLDMYHVRVDRYSAVVGQVWNNFADFATSGILRVGN  
 KXGIFRDRKPSAFLQKRWTKGMNFGSKPQGGKQ"  
 complement(6610..7326)  
 /gene="gfp"  
 /note="derived from pBBR1-GFP; expressed from gentamycin  
 resistance aacCI promoter"  
 complement(6610..7326)  
 /gene="gfp"  
 /note="modified form multi; derived from Agrobacteria victoria  
 GFP deposited in GeneBank Accession number M62653"  
 /codon\_start=1  
 /transl\_table=11  
 /product="green fluorescent protein"  
 /protein\_id="AAP04593.1"  
 /db\_xref="GI:30025560"  
 /translation="MSKGEELFTGVPIVLELDGYNKGRFSVSGEGDATTGKTL  
 KRITCTGKLVPWPLVTTFYGVQCFAPYPMHMKOHDFKAMPEGVQERTIFKD  
 DGNVKTARAEKEDGTLVNRILKGIIDFEDENILGHKLEYNSHNVITMDKONG  
 IKVNFIRNIEDSGVOLADHYQONTPIGDGVLLPDNNHYLSQSALSKDPNKRDM  
 VLEPVTAGITHGMDLYK"  
 complement(7376..8950)  
 /gene="nifA"  
 /note="derived from Klebsiella pneumoniae nifA gene of  
 pNH1; expressed from gentamycin resistance aacCI  
 promoter"  
 complement(7376..8950)  
 /gene="nifA"  
 /codon\_start=1  
 /transl\_table=11  
 /product="nif genes transcriptional activator"  
 /protein\_id="AAP04591.1"  
 /db\_xref="GI:30025558"  
 /translation="MIKSDSDPTTSPRPDLSSQFTAMORISVLSRATSEAKTIOELV  
 SYVNDARFQHEMICYDSQOETLSIEALQOETDQTLRSTQIRIRPBGGLYVLAQ  
 GOSLVLPVADDOQRFIDRLSLYDYLPIAVPLMGHSPPTGLAAHAAKORERLPA  
 CRFLFETVNLIAOTIRLMITLPTSAQAQOSPRIERPACPSRFGLENNVGSIPA  
 MROIMDIROVSRMDTVLVRSQSGKELINAIHNSPRAAFAVKENCALPNL  
 LBSLELGEKGAFTGAVROKRFELADGTLFLDITGSSASFQAKLRILIOEGME  
 RVGDETLKVVNRIIAATRRHLEBVRGHPEDGTYVRLNMPIALPRLRQEDIAE  
 LAFVLKRIASHQGRTLISDGIIRLIMYSFGNVRRENCIERAVLSSEGLIRD  
 VLFNHRDNPRLASSGPAEDGWLNSLDERORLIALLKAGWQAKARLLGMPTR  
 QVAVRIQIMDIMPRL"  
 8951..10298  
 /note="derived from pQ200KS"  
 complement(9520..9525)  
 /note="derived from gentamycin resistance aacCI gene of  
 pQ200KS"  
 complement(9548..9553)  
 /note="derived from gentamycin resistance aacCI gene of  
 pQ200KS"  
 complement(10902..11762)  
 /gene="dia"  
 /note="derived from pBBR1-GFP"  
 complement(10902..11762)  
 /gene="dia"  
 /codon\_start=1  
 /transl\_table=11  
 /product="beta-lactamase"  
 /protein\_id="AAP04592.1"  
 /db\_xref="GI:30025559"  
 /translation="MSIOHFRVALIPFAFCLPVEAPBETLVKVAEDOLGARGVY  
 ITLDNSGKILSRPREERFPMMSTFKVLGCAVLSRIDAGQOLGRRLIHSGNDLVE  
 YSPVTRKHLTDCKTYRELSSAATMSDNTANALLTTGCPRELTAFLLNMGHVRRL  
 DRWEPLEAIENDEKDTITPVAAMATTKLITGELITLASROQLIDNMEADKVAQPL  
 LNSALPWGFIIDKSGAGRGSGIILAIGPQKPSRIVIVITTSQATWDERNROIA  
 EIGASLIKHM"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,056-115  
 Score: 1718.50  
 Percent Similarity: 69.00%  
 Best Local Similarity: 54.05%  
 Length: 11973  
 Matches: 340  
 Conservative: 94  
 Mismatches: 172

Query Match: 51.24% Indels: 23  
 DB: 12 Gaps: 10  
 US-10-757-093-4 (1-634) x AY237648 (1-11973)

QY 18 GYTHrProAlAlArGHIsPheProArGAnGlu--MetThrGlnHisGluGlnProL 37  
 DB 6597 GGAATTCCTGACCGCGGGGATCCACAGAAACGATATGATTCAGTTCAGCGAAGCT 6538  
 QY 37 eu-----IleLysValArgProGlnArgThrSerSerArgGluLeuValA 52  
 DB 6537 TGGCTGCAGTCAGTCCCTTATGTCGCTCTGTAGAAACCCCAACCCGTAAATTCAAA 6478  
 QY 52 snLeuAspGlyLeuTrpLysPheAlaLeu-----AlaSerGlyLeuAsnAspThrA 69  
 DB 6477 AACTGACGCGCTGGCATTCAGTCGATCCGGAATACTGGATTCATCAG---- 6422  
 QY 69 IAGInProTrpThAlArProLeuProLysGlyLeuGluCySProValProAlaSerTyR 89  
 DB 6421 --CGTTGGGGAAGACCGCTTACAGAAACCGGCAATTCGTGTCAGGCGAGTTTAA 6364  
 QY 89 snAspIlePheIleSerArgGluIleHisAspHisValGlyTrpValTyTrpGlnArg 109  
 DB 6363 ACGATCACTTCCGCGATGCAATTCGTAATATGCGGCAACGTCGTGATCAGCGCG 6304  
 QY 109 IuValIleValProLysGlyTrpSerGlnGluArgTyRLeuValArgAlaGluSerAlaT 129  
 DB 6303 AAGCTTATATCCGAAAGGTGGGACAGCGCATTCGTCGCTTCGATTCGCGTCA 6244  
 QY 129 hriHsHsIAGIYArgIleTyRValAsnAsnArgLeuValAlaGluHisValGlyTyR 149  
 DB 6243 CTCATTACGGCAAGGTGGGTCAATATACAGAAAGATGAGCATCAGGCGGCGCTATA 6184  
 QY 149 hrProPheGluAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeuThri 169  
 DB 6183 CGGCATTGGAAGCGGATGATCAGCGCGATGTTATGCCGGGAAAGCTACGATCACCG 6124  
 QY 169 IeGlyValAsnAsnGluLeuThriSgluThriIleProProGlyLysIleThrThrGlyA 189  
 DB 6123 TTGTGTGAACAGAACTGAACCTGACGACTATCCCGCGGGAATGTGATTAAC--G 6067  
 QY 189 snAlaThrGlyLysArgIleGlnThrTyRglnHisAspPheTyRAsnTyRAlaGlyLeuA 209  
 DB 6066 ACGAAACGCGCAAGAAAGCAAGCTTACTTCATGATTTCTTAACTAGCGGATCC 6007  
 QY 209 IaaArgSerIleTrpLeuTySerValProGlnGlnHisIleGlnAspIleThrValValT 229  
 DB 6006 ATGCGACGCTAATGCTCTACACACGCGCAACCTGGGTGAGCATATACCGTGTGA 5947  
 QY 229 hrAspVal-----AspGlyAspAsnGlyLeuIleAsnTyRgluValAlaGluAlaAsnG 247  
 DB 5946 CGCATGTCGCGCAAGACTGTACACGCGCTGTGTGACTGCGAGCGT--GTGGCCAAAT- 5891  
 QY 247 IInThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAlaLysA 267  
 DB 5890 -----GGTGAATGTCAGCGCTGAAGCTGCGTATCGCATCAACAGGTGTGCACTG 5839  
 QY 267 IAserGlyAlaGlnGlyThrValThriIleProSerValLysLeuTrpGlnProGlyAlaA 287  
 DB 5838 GACAAAGCACTACCGGAGACTTTTCAAGTGTGAATCCGACCTCTGCAACCGGGGAAG 5779  
 QY 287 IeTyRLeuTyRglnLeuGlnValAsnIleValGlySerSerGlyAspValValAspThrT 307  
 DB 5778 GTTATCTTATGAACGTGCGT-----ACAGCCAAAGCCAGACAGAGTGTGATATCT 5725  
 QY 307 yzAsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIleAsnG 327  
 DB 5724 ACCCGCTTGGCGCGCATCCGCTCAGTGCAGTGAAGGCGCAACAGCTTCCTGATTAAC 5665  
 QY 327 IyLysProPheTyRPhetThrGlyPheGlyLysHisGluAspThrAlaValArgGlyLysG 347  
 DB 5664 ACAAACCGTTACTTACTGCTTGTGTCATGATGCGGAGCTTACGTGCGCAAG 5605

QY 347 IyHisAspProAlaTyRMetValHisAspPheGlnLeuMetLysTrpIleGlyAlaAsn 367  
 DB 5604 GATTTCGATAACGTGCTGATGTCACAGCACCATTAATGACATTCGATTTGGGGCAACT 5545  
 QY 367 erPheArgThrSerHisGlyTyRProTyRAlaGlnGluValMetAspPheAlaAspArgAsnG 387  
 DB 5544 CTTACCGTACTGCTGATTAACCTTACGCTGMAAGATGCTCGACATGGGCAATGAACATG 5485  
 QY 387 IyIleValValIleAspGluThrProAlaValAlaGlyLeuAsnIleAlaLeu--MetGlyV 406  
 DB 5484 GCATCGGTGATGATGATAAATGCTGTGTGGCTTTAACTCTCTTTAAGGATGTGTT 5425  
 QY 406 alSerGluSerGlyAlaPro--GlnThrPheThrProAspAlaIleAsnAspLysThG 425  
 DB 5424 TCAGACGGGCAACAGCCGAAGAACTGTACGCAAGAGCAAGCAAGCAACCGGAAATCTC 5365  
 QY 425 IInGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHisAlaSerY 445  
 DB 5364 AGCAACGCGACTTACAGCGGATTAAGACTGATACGCGTGAACAAACCAACCAACGCG 5305  
 QY 445 alValMetTrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArgGluTyR 465  
 DB 5304 TGGTGAATGAGATATGTCGAACGAACCGATATCCGTCGCAAGTGCACGGAATATT 5245  
 QY 465 heGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaA 485  
 DB 5244 TCGCGCCACTGGCGGAAGCAAGCGTAACCTGACCCGACGCGTCCGATCACCTGCGTCA 5185  
 QY 485 snValGlyThrAlaThrTyRglnLeuAspArgIleSerAspLeuPheAspValSerCyS 505  
 DB 5184 ATGTAATGTTTGGCAGCGTCAACCGATACATCAGGATCTCTTATGTGCTGTCC 5125  
 QY 505 IeAsnArgTyRheGlyTrpTyRTrpTySerGlnThrGlyAspLeuGluValaGluAlaL 525  
 DB 5124 TGAACCTTATTAAGATGATGATATGTCAAAGCGGATTTGGAAACGACAGAAAGTAC 5065  
 QY 525 euGluLysGluLeuLeuIleGlyTrpGlnGluLysPheHisArgProIleValMetThrGluT 545  
 DB 5064 TGGAAAAAGAACTTCTGGCGCTGCGAGAAACGATCAGCCGATTTTCATCACCGAAT 5005  
 QY 545 yRgIyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGluLys 565  
 DB 5004 ACGGGGTGATACGTTTACCGGGCTGCACTCAATGTACACGACATGGAGTGAAGGT 4945  
 QY 565 heGlnValGlnMetLeuAspMetTyRHisArgValPheAspArgIleGluSerMetAlaG 585  
 DB 4944 ATCATGTGTCATGCTGATATGATATCAACCGCGCTTGTGATCGCGCGCGCTCTCG 4885  
 QY 585 IyGlnHisValTrpAsnPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAspG 605  
 DB 4884 GTGAACAGATATGGAATTTCCCGCATTTTGGCACTCGCAAGGATATTCGGCGTTGGCG 4825  
 QY 605 IyAsnLysLysGlyValaPheThrArgAspArgTyRProLysAlaAlaHisSerLeuA 625  
 DB 4824 GTAAACAAAGGAGATCTTCACTCGCACCGCAACGAAAGTCCGCGCTTTCTGCTGC 4765  
 QY 625 rGAlaArgTrpThrSerIleAsp 632  
 DB 4764 AAAAAGCTGACCTGGCATGAAC 4742

RESULT 8  
 AF305918 2103 bp DNA linear BCT 19-SEP-2001  
 LOCUS Escherichia coli O157:H- beta-glucuronidase (uidA) gene, complete  
 DEFINITION cde.  
 ACCESSION AF305918  
 VERSION AF305918.1 GI:13194573  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Escherichia coli  
 Escherichia coli  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Escherichia.  
 REFERENCE  
 1 (bases 1 to 2103)



QY 478 ThrArgProIleThrPheAlaAsnValIGlyThrAlaThrTyrGlnLeuAspArgIleSer 497  
 DB 1490 AGCGCGCCATCATCCGCTCATATGATGTTCTGCGACGCTCACACCCATCCATCAAC 1549  
 QY 498 AspLeuPheAspValSerCysIleAsnArgTyrPheGlyTyrPheSerGlnThrGlyAsp 517  
 DB 1550 GATCTCTTGATGTGTGTGCTGCGTCAACCTTATACGTTGGTATGTCGAAAGCGCGCAT 1609  
 QY 518 LeuGluGluAlaGluAlaLeuGluGlySerGluLeuHisGlyTyrPheGlnIleuSerPheHis 537  
 DB 1610 TTGGAAACCGCAGAGAAAGTACTGGAAAGAACTTCTGGCTGGCAGAGAAAGCTGCAT 1669  
 QY 538 ArgProIleValMetThrGlnTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeu 557  
 DB 1670 CAGCCATTATATCATCCGAAATACGGCTGAGTACCTTAGCCGGCTGCATCAATGTC 1729  
 QY 558 GlyLeuProTyrSerGluGluPheGlnValGlnMetLeuAspMetTyrHisArgValPhe 577  
 DB 1730 ACCGACATGTGAGTGAAGATATCATGTGATGCTGATGATATGATATCACCGCGCTTT 1789  
 QY 578 AspArgIleGluSerMetAlaGlyGluHisValTyrAsnPheAlaAspPheGlnThrAsn 597  
 DB 1790 GATCGCGTACAGCCCTCGCTGCGTGAACAGTATGAAATTTCCCGATTTTGGCATCTGG 1849  
 QY 598 LeuGlyIleLeuArgValAspGlyAsnLysGlyValPheThrArgAspArgLysPro 617  
 DB 1850 CAAGGCAATGTGGCGCTTGGCGGTACAGAGAGGAGATCTTACCCGCGACCGCAACCG 1909  
 QY 618 LysAlaIleAlaHisSerLeuArgAlaArgTyrPheSerIleAsp 632  
 DB 1910 AAGTCGCGCGCTTTCTGCTGCMAAAACGCTGACATGCATGAC 1954

RESULT 9  
 AX025617 4299 bp DNA linear PAT 16-SEP-2000  
 LOCUS AX025617  
 DEFINITION Sequence 17 from Patent WO029592.  
 ACCESSION AX025617  
 VERSION AX025617.1 GI:10187285  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 'synthetic construct  
 'synthetic construct  
 other sequences; artificial sequences.

REFERENCE  
 1 Logemann, E., Somssich, I., Hahlbrock, K., Kirsch, C. and Rushton, P.  
 TITLE Chimeric promoters capable of mediating gene expression in plants  
 JOURNAL upon pathogen infection and uses thereof  
 Patent: WO 0029592-A 17 25-MAY-2000;  
 MAX PLANCK GEBILDSCHAFT (DE) ; LOGEMANN ELKE (DE) ; SOMSSICH IMRE  
 (DE) ; HAHNBROCK KLAUS (DE) ; KIRSCH CHRISTOPH (DE) ; RUSHTON PAUL  
 (GB)

FEATURES  
 source location/Qualifiers  
 1..4299  
 /organism="synthetic construct"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:32630"  
 /note="synthetic, no natural origin"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 6,98e-116 Length: 4299  
 Score: 1717.00 Matches: 332  
 Percent Similarity: 70.02% Conservative: 93  
 Best Local Similarity: 54.70% Mismatches: 166  
 Query Match: 51.19% Indels: 16  
 DB: Gaps: 9

US-10-757-093-4 (1-634) x AX025617 (1-4299)

QY 33 HisGluGlnProLeuIleLysValArgProGlnArgThrSerSerArgGluLeuValAsn 52  
 DB 2103 CAGCGCTGAGTGGCCACCATGTCGTCCTGTAGAAACCCCAACCCGTGAATCAAAAAA 2162  
 QY 53 LeuAspGlyLeuTyrLysPheAlaLeu-----AlaSerGlyLeuAsnAspThrAla 69

DB 2163 CTCGACGCGCTGTGGCATTCAGTCTGATGCGCAAAACGTGGAAATGATCAG----- 2216  
 QY 70 GlnProTyrThrAlaProLeuProLysGlyLeuGluCysProValProAlaSerTyrAsn 89  
 DB 2217 CGTTGGTGGAAAGCGCCCTTACAGAAAGCCGGGCAATTCCTGTGCGCAGGAGTTTAAAC 2276  
 QY 90 AspIlePheIleSerArgGlnIleHisAspHisValIGlyTyrValTyrTyrGlnArgGlu 109  
 DB 2277 GATCAGTTCGCCGATGCAGATATTTGTATTTATGCGGGCAACAGTGTGATACGCCGAA 2336  
 QY 110 ValIleValProLysGlyTyrSerGlnLysArgTyrLeuValArgAlaGluSerAlaThr 129  
 DB 2337 GTCTTATACGAAAGTGTGGCAGAGCGGATATGCGTTCGCTTCGATCGCGCTACT 2396  
 QY 130 HisHisGlyArgIleTyrValAlaAsnAspArgLeuValAlaGluHisValIGlyTyrThr 149  
 DB 2397 CATTCAGCGCAAGTGTGGTCAATATATCAAGAAAGATGAGCATCAGAGGCGGCTATACG 2456  
 QY 150 ProPheGluAlaAspValThrGluLeuValAlaProGlyGluLysPheArgLeuThrIle 169  
 DB 2457 CATTTGAAGCCGATGTACAGCCGATATTTATGCGGGGAAAGTACGATATCACCGTT 2516  
 QY 170 GlyValAsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThrGluAsn 189  
 DB 2517 TGTGTGAACACGAACTGAATGCGACATACCCGCGGAGATGTGATTAC--GAC 2573  
 QY 190 AlaThrGlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAla 209  
 DB 2574 GAAACCGGCAAAAGAGCTTACTTCAATGATTTCTTAACTTATACCGGAAATTCAT 2633  
 QY 210 ArgSerIleTyrLeuTyrSerValProGlnGlnHisIleGlnAspIleThrValValThr 229  
 DB 2634 CCGACGGTATATCTTACACACCGCGCAACACTGGGAGAGATATCACCGTGTGACG 2693  
 QY 230 AspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGluValIGluValAlaAsnGln 247  
 DB 2694 CATGTGCGCAAGACTGTATACACACCGCTGTGTGATGCTGACAGGTG--GTGCCAAT-- 2747  
 QY 248 ThrThrGlyGlnIleGlnIleSerValIleAspGluAspGlyAlaIleValAlaLysAla 267  
 DB 2748 -----GGGATGTACGCGTTAACTGCGCTGATGCGGATCAACAGGTGTTCACATGGA 2801  
 QY 268 SerGlyAlaGlnGlyThrValThrIleProSerValLysLeuTyrPheProGlyAlaAla 287  
 DB 2802 CAAGGCACTAGGGGAGCTTGCAGAGTGATGCCACCTCTGGCAACGGGTGAAGGT 2861  
 QY 288 TyrLeuTyrGlnLeuGlnValAsnIleValIGlySerSerGlyAspValValAspThrTyr 307  
 DB 2862 TATCTCTATGAACGTGCGCTC-----ACAGCCAAAGCCAGACAGAGTGTATATCTAC 2915  
 QY 308 AsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIleAsnGly 327  
 DB 2916 CCGCTTCGCGTGGGATCCGCTGACGTGAGTGAGTGAAGGGGAAACAGTCTGATTAAACAC 2975  
 QY 328 LysProPheTyrPheThrGlyPheGlyLysHisGluAspThrAlaValArgLysGly 347  
 DB 2976 AAACGCTTACTTACTGCTGCTTGGTGTGCTCATGAAGATCCGACATTCGTGCTGCAAGGA 3035  
 QY 348 HisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTyrIleGlyAlaAsnSer 367  
 DB 3036 TTGGAATACGTGCTGATGTGACAGCAACGATTAATGACTGGATGGGGCCAACTGC 3095  
 QY 368 PheArgThrSerHisTyrProTyrAlaGluGluValMetAspPheAlaAspArgAsnGly 387  
 DB 3096 TACCGTACTCGCATTAACCTTACGCTGAAGATGCTGACCTGGCAGATGAATGATGCG 3155  
 QY 388 IleValIleAlaAspGluThrProAlaValIGlyLeuAsnIleAlaLeu---MetGlyVal 406  
 DB 3156 ATCGTGATATTGATGAACGTGCTGTGCGCTTAACTCTTAAAGCAATGTGTTTC 3215  
 QY 407 SerGluSerGlyAlaPro---GlnThrPheThrProAspAlaIleAsnAspLysThrGln 425

Db 3216 GAAGCGGCGAACAAGCCGAAAGTGTACAGCGAAGAGGAGGAGTCAACGGGGAACCTCAG 3275  
 Qy 426 GUAAlaHelyeGlnAlaAargGluLeuAlaAargAlaAargAlaAargAlaAargAlaAargVal 445  
 Db 3276 CAAGCCACTTACAGCGATTAAGAGCTGATAGCGGTACAAACCAACCAAGCGTG 3335  
 Qy 446 ValMetTrpSerIleAlaAengLupProAlaSerHisGluAspGluAlaAargGluTyPhe 465  
 Db 3336 GTGATGTGAGATTTTCGCAACGAAACCGGATACCCGTCGCAAGGTGACCGGAATATTTTC 3395  
 Qy 466 GluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsn 485  
 Db 3396 GCCCCACTGCGCGAACAACCGCTAAACTGCACCCGACGCGTCGATCACCTCGCTCAAT 3455  
 Qy 486 ValGlyThrAlaThrTyrglnLeuAspArgIleSerAspLeuPheAspValSerCysIle 505  
 Db 3456 GATATGTTCTGCACGCTCACACCGATACCATACGATCTTTGATGTGCTGTGCTG 3515  
 Qy 506 AsnArgTyPheGlyTyrglnSerGlnThrGlyAspLeuGluAlaGluAlaAlaLeu 525  
 Db 3516 AACCGTTATTAACGATGCTATGTCCAAGCGCGGATTTGAAACGCGAAGAAAGTACTG 3575  
 Qy 526 GluTyGlnLeuHisGlyTyrglnGluTyPheHisArgProIleValMetThrGluTy 545  
 Db 3576 GAAAAAGAACTTCTGCGCTGCGAGAAACTGCATCAGCCGATTCATCATCCGGAATAC 3635  
 Qy 546 GlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGluGluPhe 565  
 Db 3636 GCGGTGATACGTTACCGGCGCTGCACCTCAATGTACACCGACATGTGGATGAAAGATAT 3695  
 Qy 566 GlnValGlnMetLeuAspMetTyrglnAspValPheAspArgIleGluSerMetAlaGly 585  
 Db 3696 CAGTGTGCATGCTGATATGATATCATCCGCTCTTGTATGCTGCACGCGCGTCTGCT 3755  
 Qy 586 GluHisValTyrglnPheAspPheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGly 605  
 Db 3756 GAACAGATATGAAATTTGCGCATTTTGGACCTCGCAAGGCATATGTCGCTGGCGGT 3815  
 Qy 606 AsnLysLysGlyValPheThrArgAspArgLysProLysAlaAlaAlaHisSerLeuArg 625  
 Db 3816 AACCAAAAGAGGATCTTCACTCCGACCGCAACCGAAACGAAAGCGGCTTTTCTGCTGCA 3875  
 Qy 626 AlaArgTrpThrSerIleAsp 632  
 Db 3876 AACGCTGACACTGCATGAAC 3896  
 RESULT 10  
 LOCUS ATH298139 2142 bp mRNA linear PLN 25-MAR-2003  
 DEFINITION Arabidopsis thaliana mRNA for beta-glucuronidase (gus gene).  
 ACCESSION AJ298139  
 VERSION AJ298139.1 GI:14594806  
 KEYWORDS beta-glucuronidase; gus gene.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 1  
 De Greve, H., Nguyen, V., Deboek, F., Thia-Toong, L., Karimi, M. and  
 Herlaars, J. P.  
 T-DNA tagging of the translation initiation factor eIF-4A1 from  
 Arabidopsis thaliana  
 Plant Sci. 161, 685-693 (2001)  
 2 (bases 1 to 2142)  
 de Greve, H.  
 Direct Submission  
 Submitted (18-DEC-2000) De Greve H., Biology, Vrije Universiteit  
 Brussel, Genetische Virologie, Vrije Universiteit Brussel,  
 Paardenstraat 65, Sint Genesius Rode, B-1640, BELGIUM  
 Location/Qualifiers  
 1..2142  
 /organism="Arabidopsis thaliana"

/mol\_type="mRNA"  
 /strain="C24"  
 /db\_xref="taxon:3702"  
 /chromosome="3"  
 /note="transgenic line M18"  
 1..2142  
 /gene="gus"  
 85..1897  
 /gene="gus"  
 /codon\_start=1  
 /product="beta-glucuronidase"  
 /protein\_id="CAC43289.1"  
 /db\_xref="GI:14594807"  
 /db\_xref="GOA:O93VY4"  
 /db\_xref="UniProt/TREMBL:O93VY4"  
 /translation="MKRVETPTRETKLDGMAFSLDBNCGIDDRMWSALQESRA  
 IAVPSFNDQFADADIRNAGVNTYQREVFIPKMGAGRI VLFDAVTYGRKVVNQ  
 EWMHOGGYTPPEADVTPTVIAGSKSVRIIVCVNNEIMQTI PGWVITDENGKKQSY  
 FDFPFYAGIHSVMLYTPPTWVDITVTVHVAQDCNHSASVDMQVAVANGDVSELD  
 ADQVVATGQSGTLQVNPPLMORGEVLYELCVTAKSQTECDIPLRVGIRSVAV  
 KEGRLINRKPFTFGEGHEDADLRGKFPDVTMTHDALMDMIGANSRTSHPYA  
 KEMLDADHGIVLIDETAVGFNLSLIGFEGANKPKLISEAVNGETQOAHDAI  
 KELIARDKNHPSVWMSIANEPDTPQARERFAPLAETRKIDPTPTTCVNVFCD  
 AHTDITSDLPDYLCNRYGYGVOSGDLTAELVLEKELAMOEKIHOP1ITREYGD  
 TLAGLSMTDMWSEYOCAMLDMYRVRDRAVAVVGQVWNPADPATSGILRVGN  
 KGIPTRDKKPSAFLDLKRWTKMNFGRKPGQCGKQ"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 3,34e-116 Length: 2142  
 Score: 1716.00 Matches: 331  
 Percent Similarity: 70.43% Conservative: 93  
 Best Local Similarity: 54.98% Mismatches: 162  
 Query Match: 51.16% Indels: 16  
 DB: 8 Gaps: 9  
 US-10-757-093-4 (1-634) x ATH298139 (1-2142)  
 Qy 38 IleLysValArgProGlnAlaGlnThrSerArgGluLeuValAsnLeuAspGlyLeuTrp 57  
 Db 83 GTGATGTCGCTCTTCTAGAAACCCCAACCGGAAATCAAAAACCTCGACGCGCTGTG 142  
 Qy 58 LysPheAlaLeu-----AlaSerGlyLeuAsnAspThrAlaGlnProTrpAla 74  
 Db 143 GCATTGATGCTGATGCCGAAACTGTGATATGATGAC-----CGTTCGTGGAAAGC 136  
 Qy 75 ProLeuProLysGlyLeuGluCysProValProAlaSerTyrglnAsnAspIleSer 94  
 Db 197 GCGTTACAAAGAAAGCGCGCAATGCTGTGCGCAGGAGTTTAAACATCATCGTGGCGAT 256  
 Qy 95 ArgGluIleHisAspHisValGlyTyrglnValTyrglnAlaArgGluValIleAlaProLys 114  
 Db 257 GCAGATATTCGATATTAATGCGCAACGCTGTGATACAGCGCAACGCTTTATACCGAA 316  
 Qy 115 GlyTrpSerGlnGluArgTyrglnValArgAlaGluSerAlaThrHisIleGlyArgIle 134  
 Db 317 GGTGTGGACAGCGCGAGATTCGCTGCGCTTTCGATGCGCTCATATTAACGCAAGTG 376  
 Qy 135 TyrValAsnAsnArgLeuValAlaGluHisValGlyTyrglnThrProPheGluAlaAsp 154  
 Db 377 TGGTCATATATACAGAAAGATGAGCATCAGGCGGCGATATCGCATTTGAAGCGAT 436  
 Qy 155 ValThrGluLeuValAlaProGlyGluTyPheArgLeuThrIleGlyValAsnAsnGlu 174  
 Db 437 GTCACCGCGATGTTATGTCGCGGAAAGGTGATGATACCGCTTTGTGAAACGAA 496  
 Qy 175 LeuThrHisGlyThrIleProProGlyLysIleThrThrGlyAsnAlaThrGlyLysArg 194  
 Db 497 CTGAACGTGACAGCTATCCCGCGGAAATGCGATTAAC---GACGAAACGCGCAAGAA 553  
 Qy 195 IleGlnThrTyrglnHisAspPheTyrglnTyrglnAlaGluAlaArgSerIleTrpLeu 214  
 Db 554 AAGCACTCTTATCTTCATGATTTCTTAACTATGCGCGAATCATCATGCGAGCGTAATGCTC 613



```

/product="translation initiation factor eif-4a1"
/protein_id="CAC43286.1"
/db_xref="GI:14594802"
/db_xref="GOA:Q94CM1"
/db_xref="UniProt/TrEMBL:Q94CM1"
/translation="MGLQENLRLGIYAGFEKPSAIQKGIYVFCFKGLDVIQQAGSGT
GVTATGSGVLIQQLDPSLIQCALVLAPRELAIQLEKWRALADYLVGVHACVGT
SVREDRITLQAGHYVVGTRGVFDMKXQSLRADIKNPVLDEADEMLSRGKDOIY
DIPQLPKIQGVSPATMPREALETTRKPKSRPAIILYKREDTLEGIKQRYNVEK
EEMKLETTLDLVEYLAITQSVIFVNTNRKVDLTDMRSRDHVSATGDMQDNTDI
IKMFFSGSRVLITTDLAGIGIDVOQVSLINFDLPQEPENLARIIGSRGGRKGV
AINEFTRDDERMIFDIQKFNVVVEELPSNVADLL"
1974..2063
/gene="eif-4a1"
/number=3
2064..2494
/gene="eif-4a1"
/number=4
2495..2575
/gene="eif-4a1"
/number=4
2576..3532
/gene="eif-4a1"
/number=5
3872..3563
/gene="gus"
3872..3563
/gene="gus"
/codon_start=1
/product="beta-glucuronidase"
/protein_id="CAC43287.1"
/db_xref="GI:14594803"
/db_xref="GOA:Q93VY4"
/db_xref="UniProt/TrEMBL:Q93VY4"
/translation="MVRPEVETPREIKKLDGLMAFSLIDRENCIGIDRWESALQESRA
IAPGSPNQFADADIRNAYAGNVYOREVPIPKMAGORIVLFPDAVTHGKVMVNO
EWEHQGGYTPFADVPYIAGKSVRTVCVNNELMORTIPGAVITBENGKKOSY
FHDREYAGIHSVMLKTTNPVMDITVYVADONASVDMOVAVNDVSVERLD
ADQVVAEGGTSGLQVNPVPHMOPGEGVLYELCTYAKSOTECDIPLRIVGRSVYA
KGEQFLINRPFYFTGFRHEDLDLKGKFDNVLWYDHALMDWIGANSRTRSHYVA
EEMLDWADSHGIVVIDETAAGFNLISLGIFEAGNPKELYSSEAANGVGTQVHLCQI
KELIARDKNHPSVVMMSIANEPDTRPOGAEFAPALAEATRKLDPTPRITCOVHMC
AHTDITSDLEFVLCINRYGVMYQSGDLEAEKVLKELLAMQEKHOPITIEYGV
TLAGHSMTYDMSSEYQCAKMLDMYHRVPRVSAVVGQVWNPADPAISQGIIRVGN
KGIETRDREKPSAFLDQKMTGMFGKPKQGGKQ"
ORIGIN

Alignment Scores:
Pred. No.: 2,98e-115 Length: 11509
Score: 1716.00 Matches: 331
Percent Similarity: 70.43% Conservative: 93
Best Local Similarity: 54.98% Mismatches: 162
Query Match: 51.16% Indels: 16
DB: Gaps: 9

US-10-757-093-4 (1-634) x ATH298137 (1-11509)
QY 38 lIeLyVaIaRgProGlnArGThrSeSaRgGluLeuVaIaSnLeuAaPglYleuTrp 57
Db 3869 GTCATGCTCGCTCTGTGAAGAACCCCAACCGTGAATCAAAAAAGTCGACGGCTGTGG 3928
QY 58 lYsPheAlaLeu-----AlaSerGlyLeuAaSnAaPThrAlaGlnPrcTTPThra 74
Db 3929 GCATTAGCTGTGATCGCGAAACCTGTGAATTGATAG-----CGTTGGTGGAAAG 3982
QY 75 ProLeuProLySgLyLeuGluCyProVaIProAlaSerTyRhaAaPrllePheIleSer 94
Db 3983 GCCTTACAGAAACCGCGCAATTGCTGTCCAGCGATTTAACATCACTGCGCGAT 4042
QY 95 ArgGluIlleHisAaPheValaGlyTrpValaTyTrGlnaRgGluVaIleValProLyS 114
Db 4043 GCAGATATTGTGAATATATGCGGCAACGCTGTGATATACGCGCAAGCTTTATACGAAA 4102

```

```

QY 115 GlyTrpSerGlnGluArGTrYleuVaIaRgAlaGluSerAlaThrHisIleGlyArgIle 134
Db 4103 GGTGGGCAAGCCAGCGATATCGCGCTTTCGATCGCGTCACTTATACGCAAGTG 4162
QY 135 TrpValaAaSnAaRgLeuVaIaGluHisValaGlyGlyTrpProPheGluAaP 154
Db 4163 TGGGCAATATATCAAGAAATGATGAGCATCAGCGCGGCTATCGCATTTGAAAGCGGAT 4222
QY 155 ValThrGluLeuVaIaAlaProGlyGluLySPhaArgLeuThrIleGlyAaSnAaGlu 174
Db 4223 GTCACCCCGGTATGTTATTCGCGGAAAGTGAAGTACGATCCGTTGTGAAACAGAA 4282
QY 175 LeuThrHisGluThrIleProProGlyLySleThrThrGlyAaSnAlaThrGlyLySArg 194
Db 4283 CTGAAGTCGACAGCTATCCCGCGGAAATGATGATTAAC---GACGAAAACGGCAAGAA 4339
QY 195 lIeGlnThrTyGlnHisAaPheTyRaaTyRAlaGlyLeuAlaRgSerIleTrpLeu 214
Db 4340 AAGCAGTCTTACTTCCATGATTTCTTAATCATGCGGAATCCATCGACGATGATGCTC 4399
QY 215 TyrSerValProGlnGlnHisIleGlnAaPrlleThValaValThrAaPVal-----AaP 232
Db 4400 TACACCAACCCGCAACCTGGGTGAGCATATCACCGTGTGACGATGCGGCAAGAC 4459
QY 233 GlYAspAaSnGlyLeuIleAaSnTyRgIleVaIaGluVaIaAaSnGlnThrThrGlyGlnIle 252
Db 4460 TGTAAACACGCGCTCTGTGACTGCGAGTG--GTGACCAAT-----GTTGATGTC 4507
QY 253 GlnIleSerValIleAaPgluAaPglYAlaIleValaIleValaAaSerGlyAlaGlnGly 272
Db 4508 AGCGTTGAACGTGCGTATCGGATCAACAGTGTGCAACGTGACCAAGGCACTACGCGG 4567
QY 273 ThrValThrIleProSerValLySleuTrpGlnProGlyAlaAlaTyRleuTyRglnLeu 292
Db 4568 ACTTTCAGATGTGTGATCCGCACTCTGCGCAACCGGTAAGTTATCTATGAACTG 4627
QY 293 GlnValaAaSnIleValaGlySerSerGlyAaPValaIaAaPThrTyRhaAaSnAlaThrGly 312
Db 4628 TGGCTC-----ACAGCCAAAGCCAGACAGATGATATCTAACCCGCTTCGCGTCGCG 4681
QY 313 ValaRgTrpValaIleValaAaGlySerGlnPheLeuIleAaSnGlyLySProPheTyR 332
Db 4682 ATCCGTCAGTGCAGTGAAGGCGCAACGATCTCTATTAACCAACACGCTTACTTT 4741
QY 333 ThrGlyPheGlyLySHisGluAaPThrAlaValaRgLyLySgIlyHisAaPProAlaTyR 352
Db 4742 ACTGCTTTGCTGCTCATGAAGATGCGGACTTACGTGGCAAGATTGATACGTCGCTG 4801
QY 353 MetValHisAaPheGlnLeuMetLySTrpIleGlyAlaAaSerPheArgThrSerHis 372
Db 4802 ATGCTGACGACACGCAATTAATGACTGATTTGGGCCAACTCTTACCTACTCGCAT 4861
QY 373 TyrProTyRAlaGluGluVaIaMetAaPheAlaAaPArgAaSnGlyIleValaValIleAaP 392
Db 4862 TACCTTACGCTGAAGAGATGCTCGACTGGCGAGATGAACATGCACTCGTGATGAT 4921
QY 393 GluThrProAlaValaGlyLeuAaSnIleAlaLeu---MetGlyValaSerGluSerGlyAla 411
Db 4922 GAAATGCTGCTGCTCGCTTTAACTCTCTTTAGGATGATGTTTCGAAGCGGCAACAG 4981
QY 412 Pro---GlnThrPheThrProAaPAlaIleAaAaPlyleThrGlnGluAlaHisIleGln 430
Db 4982 CCGAAAGAACTGTACAGCGAAGGCAAGCTACAGGGGAAACTGACGACGCACTTACAG 5041
QY 431 AlaIleArgGluLeuIleAlaArgAaPlySaaHisAlaSerValaIleMetTrpSerIle 450
Db 5042 GCATTAAAGAGCTGATAGGCGGTGCAAAACACCAACGCTGGATGATGAGATAT 5101
QY 451 AlaAaSnGluProAlaSerHisGluAaPglYAlaArgGlyTyRPhaGluProLeuThrAa 470
Db 5102 GCCAACGAACCGGATACCGTCCGCAAGGTCACGGGAATATTTCCGCGCACGCGGCA 5161
QY 471 LeuThrArgGlnLeuAaPProThrArgProIleThrPheAlaAaSnValaGlyThrAlaThr 490

```



```

Db      5162 GCAACGCGTAAACTGACCCGACGGCTCGATCACCCTGCGTATGATGTTCTGCAC 5221
Qy      491 Trrglnleuasparglleserleupheaspvalsercygellaspnarglyrphgily 510
Db      5222 GCTCACACCGAATACCATTCGATCTCTTGTATGTGCTGTGCTGACCGTATATACCGA 5281
Qy      511 Trrpyserglnthrghyaspheugluaglualalaleuglnlyleuglnleuhs 530
Db      5282 TGTATGTCTCAAGAGGGGCGATTTGGAAACCGAGAAAGTACTGAAAGAAAGAACTTGTG 5341
Qy      531 GltTrrglnleuasphehisargproilleuametrnglyrlyalaaspthrleu 550
Db      5342 GCGTGGCAGAGAACTGATCAGCGATATCATACCGAATACGGCGGTGATACGTTA 5401
Qy      551 AlaglyleuhsaserlleuuglyleuprotTrrpsergluuphegnlnvalglmeleu 570
Db      5402 GCGGGGCTGCTCATATGACACCGACATGTGGAGTGAAGAGTACGTGTGATGCGCTG 5461
Qy      571 AspmetTrrhisargvalpheasparglleugusermetalagllyluhilevalTrrasn 590
Db      5462 GATATGTATCACCGCGCTTGTGATCGCGTACGCGCGTCCGTGAACAGTATGGAAT 5521
Qy      591 PhealasphegnlnthrAsnleuglylleleargvalaspgllyasnlylyeglyval 610
Db      5522 TTCCGCGATTTGGACCTCGCAAGCATATTGCGCGGTAAAGAAAGGATC 5581
Qy      611 PheThrArgAspArglyProlyAlaAlaAlaHisSerleuArgAlaArgTrrpThrSer 630
Db      5582 TTCACTGCGCAGCCGAAACCGAAGTCGGCGCTTTCTTCTGCTGCAAAAACGCTGAGTGGC 5641
Qy      631 Ileasp 632
Db      5642 ATGAGC 5647

RESULT 12
AF502128      5667 bp      DNA      circular SYN 15-MAY-2003
LOCUS      Transient expression vector pB1221, complete sequence.
DEFINITION
AF502128      AF502128.1 GI:20454202
VERSION      Transient expression vector pB1221
KEYWORDS      other sequences; artificial sequences; vectors.
SOURCE      1 (bases 1 to 5667)
ORGANISM      Chen, P.Y., Wang, C.K., Soong, S.C. and To, K.Y.
REFERENCE      Complete sequence of the binary vector pB121 and its application
AUTHORS      in cloning T-DNA insertion from transgenic plants
TITLE      Mol. Breed. 11, 287-293 (2003)
JOURNAL      2 (bases 1 to 5667)
AUTHORS      Chen, P.Y., Wang, C.K. and To, K.Y.
TITLE      Direct Submission
JOURNAL      Submitted (12-APR-2002) Institute of Bioagricultural Sciences,
AUTHORS      Academia Sinica, Taipei 11529, Taiwan
FEATURES
source
1..5667
/organism="Transient expression vector pB1221"
/mol_type="genomic DNA"
/db_xref="taxon:193296"
/note="HindIII-EcoRI fragment (3032 bp) carrying CamV 35S
promoter, gus coding region and nos terminator was
digested from the binary vector pB121 and ligated with
cloning vector pUC19, which was previously digested with
HindIII and EcoRI, binary vector pB121 sequence is
deposited in GenBank Accession Number AF485783; plasmids
pB121 and pB122 were purchased from Clontech
Laboratories, Palo Alto, CA, USA"
25..859
/note="CamV 35S promoter"
gene      896..2707
CDS      /gene="gusa"
896..2707
896..2707

```

```

/term="gusa"
/codon_start=1
/transl_table=11
/product="beta-glucuronidase"
/protein_id="AA022211.1"
/db_xref="GI:20454203"
/translaction="MURPVTPTREIKKLDGMARSLDRENCIGIDQRMWESALQESRA
IAPGSPNDQFADADIRNAGVWVQREVFIPKNAQGRIVLRPAATVYGVWVNO
EWEHOGGYTPFEADVTPEYVLAGSVRIIVCVNNEIMQTIIPGAVITDENGKXOSY
FHDFFNVAIGHRSVMLYTPNTWVDITVTHVADCNHSAVDMQVAVNAGVSLRD
ADQOVVATGGTSGTLQOVNPHLMQREGYVLEICVTAKSOTECITVLRVGRIVAV
KEBQFLNKPFPYFTGEGHNEADLRKGFVNLVMDHMDLMDTIGANSYRSHYPA
KEBLADKNEHPSVWMSINBEDTRPQGRAREFALAEATRLDPTTRPIVCVWAFCD
AHTDITSDLPVLCNRYGVWVQSDLETAKEVLEKELEMAQELHQP1ITTEYGD
TLAGHSWYTDLMSESEYOCAMDHYHVRDVSAYVGVQVWVFADFATSGILRYGN
KKGIPTDRDKPSAFLLOKMTGNMFPKPGQGGKQ"
2778..3030
/note="nos terminator"
3634..4494
/gene="amp"
3634..4494
/gene="amp"
3634..4494
/codon_start=1
/transl_table=11
/product="beta-lactamase"
/protein_id="AA022212.1"
/db_xref="GI:20454204"
/translaction="MSIOHFRVALIPFPAFLCPVFAHPETLVKVAEDQLAGRVG
IELDLSKILLESFRRPEFRPMWSTFKALGAVLSRIDAGBOLGRIHNSQNDIYE
YSPVTEKHLDTGMTRVELCSAITSNDTANLLITLTGKREKLTALPLHNGDVYTL
DRMBELNEALPNDERDTPMPVAMATTIRKLTLELTLASRQQLDIMEADKRVAGL
LRSLAPACMFADKSGAGERSKGIITALGPDGKPSRIVLYITGSAITMDERNROIA
EIGASLIRKM"
4642..5285
/note="pUC19 plasmid replication origin"

rep_origin
US-10-757-093-4 (1-634) x AF502128 (1-5667)

Alignment Scores:
Pred. No.: 1,296-115 Length: 5667
Score: 1715.50 Matches: 338
Percent Similarity: 68.25% Conservative: 94
Best Local Similarity: 53.40% Mismatches: 182
Query Match: 51.15% Indels: 19
DB: Gaps: 10

Qy      10 LeuSerleuAlaAlaProSerleuGlyThrProAlaAlaArgHisPheProArgAsnGlu 29
Db      800 CTATCTTCGCAAGACCTCTCTATATAGAAAGTTCATTTCATTGGAGAGAACACG 859
Qy      30 MethTrglnHisGluGlnPro-----LeuIleuysValArgProGlnArgThrSer 46
Db      860 GGGAGCTTGAAGATCCCGGGGTGTCAGTCCCTTATGTATGCTCTGTAGAAACCCCA 919
Qy      47 SerArgIleuValAlaLeuAspGlyLeuTrrpHisPheAlaLeu-----AlaSer 63
Db      920 ACCCGTGAATCAAAAACCTGACGGCTGTGGCATTCTGCTGATCGCGAAACCTGT 979
Qy      64 GltleuAsnAspThrAlaGlnProTrrpThrAlaProleuProlysglyleuglucyPro 83
Db      980 GGAATGTATGAG-----CGTTGTGGGAAAGCGCGTTACAAAGACCGGCAATTCGT 1033
Qy      84 ValProAlaSerTrrAsnAspIlePheIleSerArgGlnIleHisAspHisValGlyTrrp 103
Db      1034 GTGCGACGCGATTTCACATTCAGTTCGCCGATGACATATTCATATATATGCGGCGAC 1093
Qy      104 ValTrrTrrglnArgGluValIleValProlysglyTrrpserglnIleArgTrrleuVal 123
Db      1094 GTCTGTATCAGCGCGAAGTCTTATATACGAAAGTrrGGCAGCGACGCTATCGTGTG 1153
Qy      124 ArgAlaGluSerAlaThrHisGlyArgIleTrrValAsnAsnArgLeuValAlaGlu 143

```





/translation="MLRPFVETPRRIKLDGLMAFSLDRENCIGDORMWESLAJESRA  
 IAVPSRNDQPADADIRNVAQVWVYRERFIRKMAAGCIVIRPVALYTHYKQVWNO  
 EVMEHQGTPEADYTPVIAKSGVRITVCNNELNMQTTIPRQVITDENGKKQSY  
 FPDFRYAGIHRSVLMYTPNTWVDITVTVHQAQCNASVDMQVNAADVSELRD  
 ADOQVATQGTSGTLQVNPBHMQGBEYLYELCTVYASQTECIIYLRVIRSVYA  
 KEOEFLINKPFYTFGFBHEDADLAKGFNDVLMWHDALMDWIGANSYRIRSHVYA  
 EELMDADHGHIVIDETAAGPNISLIGFEGANKPKLYSEEAANGTQOAHLOAI  
 KELLARDKRPVWVMSIANEPDTPQAGAREYFAPLAEATRLDPRPTCYNMFCD  
 AHTDTISDLFDVLCNRYTGWYQSGDLTAKEVLEKELIANQKHLQPIITTEYGV  
 TLGAHSMYTDMSSEYOCAMLDMYRVDVRSVAVVEQVWNPADPATISQGLRVGN  
 KKGIFTRDKPKSAPFLQKWTGMNFGKPOGGQ"

3' UTR

misc\_feature

/note="nopaline synthase gene"

3432..5667

/note="puc18"

ORIGIN

Alignment Scores:

Pred. No.:	1,29e-115	Length:	5667
Score:	1715.50	Matches:	338
Percent Similarity:	68.258	Conservative:	94
Best Local Similarity:	53.408	Mismatches:	182
Query Match:	51.158	Indels:	19
DB:	12	Gaps:	10

US-10-757-093-4 (1-634) x ECOL14112 (1-5667)

QY 10 LeuSerLeuAlaIaIaProSerLeuGlyThrProAlaIaIaArgHisPheProArgAsnGlu 29  
 DB 1198 CTAATCCTCGCAAGACCTTCTCTATATTAAGAAAGTTCATTCATTTCAGAGAACACAG 1257  
 QY 30 MetThrGlnHisGluGlnPro-----LeuIleValValArgProGlnArgThrSer 46  
 DB 1258 GGGGACTTAGAGAGATCCCGGGGTGTCAGTCCCTATGTTACGTCCTGAGAAACCCA 1317  
 QY 47 SerArgGluLeuValAsnLeuAspGlyLeuThrLysPheAlaLeu-----AlaSer 63  
 DB 1318 ACCCGGAAATCAAAAACCTCGACGGCTGTGGCATTCAGCTGAGATCGCGAAACTGT 1377  
 QY 64 GlyLeuAsnAspThrAlaGlnProThrAlaProLeuProLysGlyLeuGluCysPro 83  
 DB 1378 GGAATGATGACAG-----CGTTGGTGGGAAAGCGCTTACAAAGAAAGCCGGCAATTGCT 1431  
 QY 84 ValProAlaSerLysArgAspIlePheIleSerArgGluIleHisAspHisValGlyThr 103  
 DB 1432 GTGCCAGGCGATTTCATCGATCGATCGCGATGCAAGATATTCGATATTCGGCGCAAC 1491  
 QY 104 ValTyrTyrGlnArgGluValIleValProLysGlyTyrSerGlnGluArgTyrLeuVal 123  
 DB 1492 GTCTGATATCAGCGCAAGCTTATACGAAAGTTGGGACGCGCATCGTCTG 1551  
 QY 124 ArgAlaGluSerAlaThrHisGlyArgIleTyrValAsnAsnArgLeuValAlaGlu 143  
 DB 1552 CGTTTCGATGCGGTCTCATTCATCGCGCAAGTGTGGTCAATATACGAAAGATGATGCGAG 1611  
 QY 144 HisValGlyGlyTyrThrProPheGlnAlaAspValThrGluLeuValAlaProGluGlu 163  
 DB 1612 CATCAGGGCGGTATACGCGCATTTGAAGCCGATGTCAAGCGGTATGATTGCGCGGAAA 1671  
 QY 164 LysPheArgLeuThrIleGlyValAsnAsnGluLeuThrHisGlyThrIleProProGly 183  
 DB 1672 AGGTGACGATATCACCGTTGTGTGAACAGAACTGAACGCGACATACCGCGGGA 1731  
 QY 184 LysIleThrThrGlyAsnAlaThrGlyLysArgIleGlnThrTyrGlnHisAspPheTyr 203  
 DB 1732 ATGATGATTAC---GACGAAACGCGCAAGAAAGAGATCTTACATTCGATTTCTTT 1788  
 QY 204 AsnTyrAlaGlyLeuAlaArgSerIleTyrLeuTyrSerValProGlnGlnHisIleGln 223  
 DB 1789 AACTATGCGCGAATCCATCGCACGTAATCTCTACACCAAGCCGCAACACTGGGTGAC 1848  
 QY 224 AspIleThrValValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTyrGlu 241

DB 1849 GATATCACCGTGTGACCGCATGTCCGCCAGACCTGAACCAAGCGCTGTGATCGGACG 1908  
 QY 242 ValGluValAlaAsnGlnThrThrGlyGlnIleGlnIleSerValIleAspGluAspGly 261  
 DB 1908 GGG---GGGCGCAAT-----GATATGTCACCGTTGAACCTCGTGAATCGGATCCA 1956  
 QY 262 AlaIleValAlaIaIaSerGlyAlaGlnGlyThrValThrIleProSerValLysLeu 281  
 DB 1957 CAGGTGCTTGCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 2016  
 QY 282 TTPGlnProGlyAlaIaIaTyrLeuTyrGlnLeuGlnValAsnIleValGlySerGly 301  
 DB 2017 TCGCAACCGGCGTGAAGTATCTCTATGAACTGTGCGTC-----ACAGCCAAAAGCCAG 2070  
 QY 302 AspValValAspThrTyrAsnLeuAlaThrGlyValArgThrValLysValIaGlySer 321  
 DB 2071 ACAGAGTGTATATCTACCCGCTTGCCTGGACATCGGCTGACGTGACGTGAAGGCCAA 2130  
 QY 322 GlnPheLeuIleAsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGluAspThr 341  
 DB 2131 CAGTTCCTGATTAACCAACCAACCGTTCTACTTACTGCTTGTGTGCTGATGAAGTGGC 2190  
 QY 342 AlaValArgGlyLysGlyHisAspProAlaTyrMetValHisAspPheGlnLeuMetLys 361  
 DB 2191 GACTTACGTGGCAAGGATTCGATTAACGTCGTGATGTCACAGACCAACCATTAATGAC 2250  
 QY 362 TTPIleGlyAlaAsnSerPheArgThrSerHisTyrProTyrAlaGlnGluValMetAsp 381  
 DB 2251 TGGATTTGGGCGCAACCTCCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2310  
 QY 382 PheAlaAspArgAsnGlyIleValValIleAspGluThrProAlaValGlyLeuAsnIle 401  
 DB 2311 TGGCAGATGAACATGCGATCTGTGATTAAGAACTGCTGCTGCTGCTGCTGCTGCTGCT 2370  
 QY 402 AlaLeu---MetGlyValSerGluSerGlyAlaPro---GlnThrPheThrProAspAla 419  
 DB 2371 TCTTTAGCGATGATTTTCGACCGGCGCAACCGGCAAGAAAGAACTGACCGCAAGGGA 2430  
 QY 420 IleAsnAspLysThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAsp 439  
 DB 2431 GTCAACGGGAAACTCAGCAACGCGCTTACAGGCGATTAAGAGCTGATGACGCTGAC 2490  
 QY 440 LysAsnHisAlaSerValValMetTyrSerIleAlaAsnGluProAlaSerHisGluAsp 459  
 DB 2491 AAAAACCAACCAACGCTGTGATGAGATATTCGCAAGAAACCGGATACCGTCCGCA 2550  
 QY 460 GlyAlaArgGluTyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArg 479  
 DB 2551 GGTGCAAGGGAATATTTCCGCGCACAGTGGCGGAAGCAACCGGTAACTCGAACCCGAC 2610  
 QY 480 ProIleThrPheAlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeu 499  
 DB 2611 CGATACCTGCGTCATATTAATGTTCTGCAAGCTCAGACCGATACCATACGAGATCTC 2670  
 QY 500 PheAspValSerCysIleAsnArgTyrPheGlyTyrTyrSerGlnThrGlyAspLeuGlu 519  
 DB 2671 TTGATGTCGTGCTGCTGAACCGTATTAAGATGATAGTCAAAAGCGCGATTTGGA 2730  
 QY 520 GluAlaGluAlaLeuGluLysGluLeuHisGlyTyrGlnGluLysPheHisValArgPro 539  
 DB 2731 ACAGCAGAAAGTACTGAAAAAGAACTTCGCGCTGCGCAGAGAAAGTGTGATCGCGG 2790  
 QY 540 IleValMetThrGluTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGluLeu 559  
 DB 2791 ATTATATACCGAAATACGCGGTGATGATGATGATGATGATGATGATGATGATGATGAT 2850  
 QY 560 ProTyrSerGluGluPheGlnValGlnMetLeuAspMetTyrHisValArgValPheAspArg 579  
 DB 2851 ATGTGAGTGAAGAGATATAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 2910  
 QY 580 IleGluSerMetAlaGlyGluHisValTyrAsnPheAlaAspPheGlnThrAsnLeuGly 599  
 DB 2911 GTACAGCGCGTGTGCTGTAACAGGATGGAATTTGCGCGATTTGCGACCTCGCAAGGC 2970



QY	262	AlaIleValAlaIValAsenGlyValInGlnIValThrValIleProSerValLysLeu	281
Db	6508	CAGGTGGTTGGCACTGGACAAGGCACTACGGGACTTTGGCAAGTGTAAATCCGACCTC	6567
QY	282	TrpGlnProGlyAlaAlaIleTyrlleuTyrglnIleuGlnValAsnIleValGlySerSergly	301
Db	6568	TGGCAACCGGGGGAAGGTTATCTCTATGAACGTGTCGTC-----ACAGCCAAAAGCCAG	6621
QY	302	AspValValAspThrTyrAsnLeuAlaThrGlyValArgThrValLysValAlaGlySer	321
Db	6622	ACAGAGTGTGATATCTACCCGCTTCGCGTCGCAACCGGTACAGTGGCAAGTGAAGGCCAA	6681
QY	322	GlnPheLeuIleAsnGlyLysProPheTyrPheThrIlyPheGlyLysHisGlnLysPthr	341
Db	6682	CAGTTCCTGATTAAACCAAAACCGTTCTACTTACTGCTTTGGTGGCATGAAATACGCG	6741
QY	342	AlaValAlaGlyIlySGlyYHisAspProAlaTyrMetValHisAspPheGlnLeuMetLys	361
Db	6742	GACTTGGCTGGCAAAAGATTCGATTAACGGTCGATGGTGCACAGCACGCACTTAATGAC	6801
QY	362	TrpIleGlyAlaAsnSerPheArgThrSerHisTyrProTyrAlaGlnGluValMetAsp	381
Db	6802	TGGATTGGGGCCAACTCCTACCGTCTCCGCACTTCCTTCCTGCGTGAAGAATGCTCGAC	6861
QY	382	PheAlaAspArgAsnGlyIleValValIleAspGlyThrProAlaValGlyLeuAsnIle	401
Db	6862	TGGCGAGATGAACATGGCATCGTGTGATGTAGTAACTGCTGCTGTGGCTTTAACTTC	6921
QY	402	AlaLeu--MetGlyValSerGluSerGlyAlaPro--GlnThrPheThrProAspAla	419
Db	6922	TCTTTTAGGCATTGGTTTGGAGCGGGCAACAACCGAAAGAACTGTACAGGAAGAGGCA	6981
QY	420	IleAsnAspLysThrGlnGluAlaHisLysGlnAlaIleArgGlnLeuIleAlaArgAsp	439
Db	6982	GTCAAACGGGAAACTCACCAAGCGCACTTACAGCGATTAAAGACTGATGACGGCTGAC	7041
QY	440	LysAsnHisAlaSerValValMetIrrPserIleAlaAsnGluProAlaSerHisGluLysp	459
Db	7042	AAAAACCAACCAAGGCTGTGTGTGGAGTATGGCCAAACCGGATATCCCGCTCCGCA	7101
QY	460	GlyAlaArgGluTyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArg	479
Db	7102	GGTGACCGGAATATTTGGCCGCACTGGCCGGAAGCAACGCGTAAACTGACCCGACGGT	7161
QY	480	ProIleThrPheAlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeu	499
Db	7162	CCGATCACTGGCGCAATGTATATGTTTCGCGACGCTCAACCGGATATCATGACGATCTC	7221
QY	500	PheAspValSerCysIleAsnArgTyrPheGlyTrrPtyrSerGlnThrGlyAspLeuGlu	519
Db	7222	TTTGATGTGCTGTGCTCAACCGTTATTAACGATGTATGTGCCAAAGGGGCGATTGGAA	7281
QY	520	GlnAlaGlnAlaAlaLeuGlnLysGlnLeuHisSgIlyTrpGlnGluLysPheHisArgPro	539
Db	7282	ACGGCAGAGGAAGTACTGGAAAAAACAATTCTCGGCTCGCAGAGAAACTGCATCAGCCG	7341
QY	540	IleValMetThrGluTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeu	559
Db	7342	ATTATCATCACCGAATATACGGCGTGATACGTTAGCCGGGCTGCATCATGACACCGAC	7401
QY	560	ProTrrPserGlnGluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAspArg	579
Db	7402	ATGTGTGAATGAAGATATCAAGTGTGCATAGCTCGGATATGTATACCGGGCTCTTGATGTC	7461
QY	580	IleGlnSerMetAlaGlyLysIleValIleTrpAsnPheAlaAspPheGlnThrAsnLeuGly	599
Db	7462	GTCACGCGCGTCTGCGGTGAACAGTATAGGATTTCCGCGATTTTGGCACTTCGCAAGGC	7521
QY	600	IleIleArgValAspGlyAsnLysLysGlyValPheThrArgAspArgLysProLysAla	619
Db	7522	ATATTGCGCGTTGGCGGTAAACAAGAAAGGATCTTCACTCGGACCGCAACCGGAAGTCG	7581
QY	620	AlaAlaHisSerLeuArgAlaArgTrpThrSerIleAsp	632

DB	7582	CGCGCTTTTCTGCTGCAAAACGCTGACACTGCATGAAC	7620
RESULT 15	AF485783	14758 bp	DNA
LOCUS	AF485783		
DEFINITION	Binary vector pBI121, complete sequence.		
ACCESSION	AF485783		
VERSION	AF485783.1	GI:19569222	
KEYWORDS			
SOURCE	Binary vector pBI121		
ORGANISM	Binary vector pBI121		
REFERENCE	1 (bases 1 to 14758)		
AUTHORS	Chen, P.Y., Wang, C.K., Soong, S.C. and To, K.Y.		
TITLE	Complete sequence of the binary vector pBI121 and its application in cloning T-DNA insertion from transgenic plants		
JOURNAL	Mol. Breed. 11, 287-293 (2003)		
REFERENCE	2 (bases 1 to 14758)		
AUTHORS	To, K.Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (20-FEB-2002) Institute of BioAgricultural Sciences, Academia Sinica, Taipei 11529, Taiwan		
FEATURES	Location/Qualifiers		
SOURCE	1..14758		
	/organism="Binary vector pBI121"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:189807"		
	/note="constructed using pB21 from Clontech Laboratories and Bln19 described in Genbank Accession Number U09365"		
	complement(13..796)		
	/note="similar to traF in Genbank Accession Number X5459"		
	complement(790..1168)		
	/note="ColEI ori; similar to sequence in Genbank Accession Number V00268"		
	complement(1161..2344)		
	/note="similar to tetA in Genbank Accession Number X75761"		
	complement(2454..2478)		
	/note="T-DNA right border"		
	2519..2825		
	/note="NOS"		
	2838..3632		
	/gene="nptII"		
	2838..3632		
	/gene="nptII"		
	/codon_start=1		
	/product="neomycin phosphotransferase II"		
	/protein_id="AA192039.1"		
	/db_xref="GI:19569230"		
	4022..4277		
	/note="NOS"		
	4974..5808		
	/note="CaMV 35S"		
	5845..7656		
	/gene="gusA"		
	5845..7656		
	/gene="gusA"		
	/note="GUS"		
	/codon_start=1		
	/product="beta-glucuronidase"		
	/protein_id="AA192040.1"		
	/db_xref="GI:19569231"		
	translation="MRRVETPTREIKLDELNAFLSDRENCGIDOREMBSALQSPRA IAVGGNDPFDADIRNVAGNWWYOREVPIPKMGAGQIVLRDAVTHGKWNQ EWEHGGQYDFEADVIVPYIAGKSVIITVCUNNELMMOTIPGMVITDENGKKQSY FHDENFAGIHRISVMLTTPNTWVDITIVTHNAODNHSVDQVAVGVSVELRD ADQVAVATGGTGSTLOVANNPHLMQBEGLVYELCYTRAKSQGTEDCIPLAVGIRSVAV KGEFFLNHRPFYFTGFRHEDADLRKGFNDVLMVHDALMDVIGANSTRISHYPYA		

```

BEMLDWADHEGIVVIDETAIVGFNLISGIFGEAKNPKELYSSEAVNGETOAHLOAI
KELIARDKXHSVVMWSIANEPDTPQGAEEYAPALAEARKIDPRITCVAMFCD
AHTDITSDLEPDLNRYGMYVOSGLEAEKLEKLEKLEKLEKLEKLEKLEKLEKLE
TLAGHSMTDMSEBYOCAMLDHYRHRVRSVAVIGVOWNRADATSGILRVGN
KKKIFTRDRKPKSAFLQKRTGMFGEPQGGKQ"
terminator
7127..7979
/misc_feature
/contig="NOS"
/complement(8621..8646)
/notes="T-DNA left border"
/misc_feature
/complement(9156..10198)
/notes="similar to ceta in GenBank Accession Number L13842"
/misc_feature
/complement(10199..11680)
/notes="similar to cifa in GenBank Accession Number X00713"
/misc_feature
/complement(11681..12673)
/notes="similar to NPTIII gene in GenBank Accession Number
V01547"
/misc_feature
/complement(12674..13443)
/notes="similar to transposable element ISI in GenBank
Accession Number X58999"
/misc_feature
/complement(13444..13794)
/notes="similarity to NPT III gene in GenBank Accession
Number V01547"
/misc_feature
/complement(13795..14066)
/notes="similar to kila in GenBank Accession Number M62846"
/misc_feature
/complement(14141..14758)
/notes="ori V: similar to sequence in GenBank Accession
Number M20134"
rep_origin
misc_feature
Number M20134"
ALIGNMENT Scores:
Pred. No.: 4.47e-115 Length: 14758
Score: 1715.50 Matches: 338
Percent Similarity: 68.25% Conservative: 94
Best Local Similarity: 53.40% Mismatches: 182
Query Match: 51.15% Indels: 19
Gaps: 10
US-10-757-093-4 (1-634) x AF485783 (1-14758)
QY 10 LeuSerLeuAlaAlaProSerLeuGlyThrProAlaAlaArgHisPheProAlaGln 29
DB 5749 CTATCCTTCGCAAGACCTTCCTATATAGAGAAAGTTCATTTGATGAGAAACACG 5808
QY 30 MetThrGlnHisGlnPro-----LeuIleValArgProGlnArgThrSer 46
DB 5809 GGGGACTCTAGAGAGATCCCGGGGTGCTAGCTCCCTATGTTAGCTGTGAGAAACCCA 5868
QY 47 SerArgGlnLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu-----AlaSer 63
DB 5869 ACCCGGAAATCAAAAACCTCGACGGCTGTGGCATTCAGTCTGATCGCGAAACTGT 5928
QY 64 GlyLeuAsnAspThrAlaGlnProTrpThrAlaProLeuProLysGlyLeuGlnCysPro 83
DB 5929 GGAATGATGAG-----CGTTGGTGGGAAAGCGCGTTACCAAGCCGGCAATGCT 5982
QY 84 ValProAlaSerLysTrpAsnAspLysPheIleSerArgGlnIleHisAspHisValGlyTrp 103
DB 5983 GTGCGCAGGCGATTGTAACGATCGATTCGCGATGCAATATTCGTAATTTGCGGGGAC 6042
QY 104 ValTrpGlnArgGlnValIleValProLysGlyTrpSerGlnGlnArgTrpLeuVal 123
DB 6043 GTCTGGTATCGCGCAAGCTTTATACGAAAGTTGGGACGCCGATATCGTCTG 6102
QY 124 ArgAlaGlnSerAlaThrHisHisGlyArgIleTrpValAsnAsnArgLeuValAlaGln 143
DB 6103 CGTTTCGATCGCGCTCATCTATCGCGCAAGTGGGTGATATATCAAGAAATGATGAG 6162
QY 144 HisValGlnGlyTrpThrProPheGlnAlaAspValThrGlnLeuValAlaProGlnGly 163
DB 6163 CATCAAGGCGGCTATACGCAATTTGAAGCCGATGTCAGCGGTATGTTATTCGCGGAAA 6222
QY 164 LysPheArgLeuThrIleGlyValAsnAsnGlnLeuThrHisGlnTrpIleProGly 183

```

```

DB 6223 AGTGAAGTATACCGCTTGTGTGAACAAGAACTGACGACTATCCCGCGGA 6282
QY 184 LysIleThrThrGlnValAsnAlaThrGlyLysArgIleGlnThrTrpGlnHisAspPheTrp 203
DB 6283 ATGTGATTTAC---GACGAAACCGCAAGAAAGACAGCTTACTTCCATGATTTCTTT 6339
QY 204 AsnTrpAlaGlyLeuAlaArgSerIleTrpLeuTrpSerValProGlnGlnHisIleGln 223
DB 6340 AACTATGCGGAGATCATGATGACGACGATGATGCTTACACCAAGCCGAAACACTGGGTGAC 6399
QY 224 AspIleThrValValThrAspVal-----AspGlyAspAsnGlyLeuIleAsnTrpGln 241
DB 6400 GATATACCGCTGTGACGATGTGCGCAAGACTGATACCAAGCCGCTGTGTGACTGGCAG 6459
QY 242 ValGlnValAlaAsnGlnThrThrGlnIleGlnIleSerValIleAspGlnAspGly 261
DB 6460 GTG---GTGCGCAAT-----GGTGTATGAGCGTTGAAGTGTGATGCGGATCAA 6507
QY 262 AlaIleValAlaLysAlaSerGlyValAlaGlnGlyThrValThrIleProSerValLysLeu 281
DB 6508 CAGGTGTTGCACTGACAGCAAGGCACTAGCGGACTTTCGAACTGATGATCCGCACTC 6567
QY 282 TrpGlnProGlyValAlaAlaTrpLeuTrpGlnLeuGlnValAsnIleValGlySerGly 301
DB 6568 TGCAACCGGGGTGAAGATTATCTCTATGAAGTGTGCTC-----ACAGCCAAAGCCAG 6621
QY 302 AspValValAspThrTrpAsnLeuAlaThrGlyValArgThrValLysValAlaGlySer 321
DB 6622 ACAGAGTGTATATCTACACCGCTTCGCGGATCCGTCAGTGCAGTGCAGGAGGCGGAA 6681
QY 322 GlnPheLeuIleAsnGlyLysProPheTrpPheThrGlyPheGlyLysHisGlnAspThr 341
DB 6682 CAGTTCTGATTAACACAAACCGCTTCTATCTAGCTGTGCTGTGATGAAGATCGG 6741
QY 342 AlaValArgGlyLysGlyHisAspProAlaTrpMetValHisAspPheGlnLeuMetLys 361
DB 6742 GACTGCGTGGCAAGATTCGATTAAGCTGATGATGATGATGATGATGATGATGATGATGAT 6801
QY 362 TrpIleGlyValAsnSerPheArgThrSerHisTrpProTrpAlaGlnGlnValMetAsp 381
DB 6802 TGATATGGGCGCAACCTTCACCGTACCTGATTCCTTACGCTGTGCTGATGAAGATCGG 6861
QY 382 PheAlaAspArgAsnGlyLysValIleAspGlnThrProAlaValGlyLeuAsnIle 401
DB 6862 TGGCGATGATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6921
QY 402 AlaLeu---MetGlyValSerGlnSerGlyAlaPro---GlnThrPheThrProAspAla 419
DB 6922 TCTTTAGGATTTGTTTTCGAAGCGGCAACAGCCGAAAGAACTGATGACGAAAGGCA 6981
QY 420 IleAsnAspLysThrGlnGlnAlaHisLysGlnAlaIleArgGlnLeuIleAlaArgAsp 439
DB 6982 GTCAACGGGGAACCTCAGACAGCGCACTTACAGGCATTAAGAGCTGATAGCGCGTGAC 7041
QY 440 LysAsnHisAlaSerValValMetTrpSerIleAlaAsnGlnProAlaSerHisGlnAsp 459
DB 7042 AAAAACCACCAAGCGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7101
QY 460 GlyAlaArgGlnTrpPheGlnProLeuThrAsnLeuThrArgGlnLeuAspProThrArg 479
DB 7102 GTGCAACGGGAAATTTTGGCGCACTGGCGGAAAGCAAGCGGTAACTGACCGGACGGGT 7161
QY 480 ProIleThrPheAlaAsnValGlyThrAlaThrTrpGlnLeuAspArgIleSerAspLeu 499
DB 7162 CCGATACCTGCGCTCATGTAATGTTCTGCGACGCTCAGACCCATCATCAGCGATCTC 7221
QY 500 PheAspValSerCysIleAsnArgTrpPheGlyTrpTrpSerGlnHisGlyAspLeuGln 519
DB 7222 TTGATGTGTGCTGTAACCGTTATACGATGATGATGATGATGATGATGATGATGATGATGAT 7281
QY 520 GlnAlaGlnAlaLeuGlnLysGlnLeuHisGlyTrpGlnGlnLysPheHisAspPro 539
DB 7282 ACGCAGAGAGAGTACTGGAAAAAGAACTTCTGCGCGAGGAGAACTGATCAGCGG 7341

```

```
QY 540 IleValMetThrGluTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeu 559
Db 7342 ATTATCATCACCGAATACGGCCGGATACGTTAGCCGGCTGCATCAATGTAACACCGAC 7401
QY 560 ProTyrSerGluGluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAspArg 579
Db 7402 ATGTGAGTGAAGAGTATCATGTGTCATGCTGATATGATCATCACCGCTCTTTGATCGC 7461
QY 580 IleGluSerMetAlaGlyGluHisValTyrAsnPheAlaAspPheGlnThrAsnLeuGly 599
Db 7462 GTCAGCGCCGCTCGCTCGTGAACAGTATGGAATTTGCCGATTTTGGACCTCGCAAGGC 7521
QY 600 IleIleArgValAspGlyAsnLysLysGlyValPheThrArgAspArgLysProLysAla 619
Db 7522 ATATTCCGCGCTTGGCGGTACAGAAAGGATCTTCACTCGCGAACCGCAACCGAAGTCCG 7581
QY 620 AlaAlaHisSerLeuArgAlaArgTyrThrSerIleAsp 632
Db 7582 GCGGCTTTTCTGCTGCAGAAACGCTGACTGCGCATGAAC 7620
```

Search completed: March 22, 2005, 20:24:58  
Job time : 6487 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - nucleic search, using frame\_p2n model

Run on: March 22, 2005, 18:23:41 ; Search time 4458 Seconds  
(without alignments)  
5413.357 Million cell updates/sec

Title: US-10-757-093-4  
Perfect score: 3354  
Sequence: 1 MKRLTGLSLSLAAPSIGTP.....RKPKAAHSIRARWTSIDKN 634

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

-MODE=frame\_p2n.model -DEV=xlp  
-Q=/cgr2\_1/USPTO.spool\_p/US10757093/runat\_18032005\_164456\_27770/app.query.fasta.1.775  
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi -LIST=45  
-DOCALLGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-DEV=US10757093\_@CGN\_1\_15180\_@runat\_18032005\_164456\_27770 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

EST:  
1: gb\_esc1.\*  
2: gb\_esc2.\*  
3: gb\_hic.\*  
4: gb\_esc3.\*  
5: gb\_esc4.\*  
6: gb\_esc5.\*  
7: gb\_esc6.\*  
8: gb\_ges2.\*  
9: gb\_ges2.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1267	37.8	2473	3	AY321342 Rattus no
2	1253.5	37.4	2095	6	CD014094 CD014094
3	1175	35.0	1995	6	CD014092 CD014092
4	1168.5	34.8	2274	3	AK041058 Mus muscu
5	1045.5	31.2	1853	6	CD014093 CD014093
6	946	28.2	1377	3	CRS93823 CRS93823
7	803	23.9	1124	6	CD0503076 CD0503076
8	797.5	23.8	1051	5	BX363460 BX363460
9	797	23.8	921	9	CL486845 CL486845

10	770.5	23.0	1138	6	CD503098
11	726.5	21.7	857	5	BX745933
12	712.5	21.2	817	9	CL679241
13	705.5	21.0	906	5	BO941196
14	689	20.5	1055	4	BM557676
15	683	20.4	877	7	CF406520
16	681.5	20.3	767	7	CF255373
17	674.5	20.3	740	7	CF521612
18	671.5	20.0	816	7	CK467358
19	671.5	20.0	878	7	CN155220
20	661.5	19.7	914	6	CB203472
21	656	19.6	1063	5	BX401772
22	649.5	19.4	915	5	BO678153
23	648.5	19.3	846	6	CA453907
24	647.5	19.3	802	5	BU357212
25	643	19.2	733	7	CN264047
26	637	19.0	736	5	BP454056
27	633.5	18.9	791	5	BU239978
28	630.5	18.8	689	7	CR442804
29	622.5	18.6	725	5	BU315788
30	621	18.5	566	7	CV169266
31	620.5	18.5	900	6	CA489317
32	609.5	18.2	799	5	BU366987
33	609	18.2	771	4	BG121498
34	605.5	18.1	731	7	CV105482
35	601.5	17.9	730	7	CO559438
36	601	17.9	680	7	CK836522
37	591	17.6	898	5	BU179563
38	590	17.6	962	5	BQ684422
39	587.5	17.5	773	7	CF748482
40	585.5	17.5	757	4	B1851705
41	583.5	17.4	847	5	BX621253
42	582	17.4	925	5	BO679162
43	578.5	17.2	624	4	B1445636
44	577.5	17.2	673	6	CB841568
45	573	17.1	857	6	CB559443

## ALIGNMENTS

RESULT 1  
AY321342  
LOCUS AY321342 2473 bp mRNA linear HTC 16-JUN-2003  
DEFINITION Rattus norvegicus Ac2-223 mRNA, complete cds.  
ACCESSION AY321342  
VERSION AY321342.1 GI:32527744  
KEYWORDS  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus

REFERENCE  
1 (bases 1 to 2473)  
Xu,C.S., Li,W.O., Li,Y.C., Wang,L., Wang,S.F., Han,H.P., Wang,G.P.,  
Chai,L.Q., Yuan,J.Y., Yang,K.J., Yan,H.M., Chang,C.F., Zhao,L.F.,  
Ma,H., Shi,J.B., Rahman,S., Wang,Q.N. and Zhang,J.B.  
Liver regeneration after PH  
Unpublished  
2 (bases 1 to 2473)  
Xu,C.S., Li,W.O., Li,Y.C., Wang,L., Wang,S.F., Han,H.P., Wang,G.P.,  
Chai,L.Q., Yuan,J.Y., Yang,K.J., Yan,H.M., Chang,C.F., Zhao,L.F.,  
Ma,H., Shi,J.B., Rahman,S., Wang,Q.N. and Zhang,J.B.  
Direct Submision  
Submitted (11-JUN-2003) Henan Bioengineering Key Lab, Henan Normal  
University, NO. 148 Jianshe Road, Xinxiang City, Henan 453002, P.R.  
China

TITLE  
JOURNAL  
AUTHORS  
REFERENCE  
TITLES  
JOURNAL  
FEATURES  
source  
1. 2473  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10116"  
41. 2467  
/note="liver regeneration related protein LRRG134"

CDS  
CL486845

```

/codon_start=1
/product="Ac2-223"
/protein_id="AAP6274.1"
/db_xref="GI:32527745"
/translation="MKRDKPQKQVEKYLMDLNSRVIAQNERDESEAKRPSWTAGI
PASMSPRSVCMFVYGLCSALAQGMLEPKETPSRELKVLDDGMSFPADYNNR
LQGEKQWYRQPLRESGTLMDIPVSSFNDIQTOAELENIIGWVYREAEALPQRWQ
DTRRVLRLINSAYYAVVWNGIHVHEGHLPEFADITKLQWSPILTPRTTIAI
NNTLTPYLPPTGIYKTPSPMKYPKGFVODISDFPKNAGHSVLYXTPTTYID
ITVITDVRDYGIVNWYISUGSDHFCQLEVLDBDKIYAKGTGNGQLEVPRAHLM
MPYLMHPRALYLSLEVTMTTPESVSDFTLVGIRYVAIVKSKFLNGKRFYQVNA
KHEBSDIKRGEDWPLIKDNLRLWLGANSFRSHQPYSEVYLCDRGIIVIDEC
PVGIVLPQSFQGNVSLRHLEVMDELKMPHAPVVMVSVANPVSILKPAQYFEL
IVFQGLKTSIDSPRTGTVLVNHYAPYVDVICNSYLSWYHGHLEVILOLQSFEN
WYKMYQKPIIQSEYGADAVSGHEDPMPMESEYOTALLNHYHLIDEXKREYVIGEL
IWNFADFWTNGSPLRVTKGNKGIPTLRORPKMAAFIIRERYRIANTRTGLIQVNSF
VVDDEKVENOHICDADADAKLQHKVONVGDSTIYAAGALICSTCTNRHAEKCKST
VPSPDAGAQERGOQLSIOATGMVAVTNKASHFQVROLKTHVAHSIAVERPDRPLNS
CLLDISAGSA"

ORIGIN

Alignment Scores:
Pred. No.: 1,66e-130 Length: 2473
Score: 1267.00 Matches: 266
Percent Similarity: 57.74% Conservative: 96
Best Local Similarity: 42.42% Mismatches: 193
Query Match: 37.78% Indels: 72
DB: 3 Gaps: 13

US-10-757-093-4 (1-634) x AY321342 (1-2473)

QY 42 ProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu 61
Db 269 CCGAAGAGAGACCCCTTCGGCGGAACCTCAAGGTGCTGAGCGCGCTGTGAGACTCCG 325
QY 62 AlaSerGlyLeuAsnAspThrAlaGlnPro-----TrpThrAlaProLeu 76
Db 326 GCCGATTACTCGAACCAATCCGCTTGACAGGCTTCGAGAGACAGTGGTACCGGACCGCTA 385
QY 77 ProLysGly-----LeuGluCysProValProAlaSerTrpAsnAspIlePheIle 93
Db 386 CGGAGATCGGGGCCCAACCTTGACATGCCGCTCCCTTCAGACTCAATGACATCACCCAA 445
QY 94 SerArgGluIleHisAspHisValGlyTrpValTyTrpGlnArgGluValIleValPro 113
Db 446 GAAGCAGAGGCTTGGAACCTTCATGTGGCTGGTGATGAAGCGGAGAGGCTTCCA 505
QY 114 LysGlyTrpSerGlnGlu-----ArgTrpLeuValArgAlaGluSerAlaThrHis 130
Db 506 CAGCGATGAGCCACGAGCACCGACAGAGAGTGTGTGTGAGATCAACAGCCGCTTAC 565
QY 131 HisGlyArgIleTyValAsnAsnArgLeuValAlaGluHisValGlyTyTrpThrPro 150
Db 566 TATGCAAGTTGTGGGTGATGGATTCAATGTGTGGAACATAGAGGAGGTCACTCCGCC 625
QY 151 PheGluAlaAspValThrGluLeuValAlaProGly-----GluLysPheArgLeuThr 168
Db 626 TTGGAGGCTGACATCAACCAAGCTGGTCCAGAGTGGGCCCTTCACACCTTCGGGGTACC 685
QY 169 IleGlyValAsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThr----- 186
Db 686 ATGCGCATCAACAACACTGACCCCTTATACCTTCCACCGGGAGCCATTGTCTACAAAG 745
QY 187 -----ThrGlyAsnAlaThrGlyLysArgIleGlnThrTyTrpGlnHisAspPheTy 203
Db 746 ACTGATCTTCCTCATGTATCCCAAGGGTACTTCGTCCAGAGACATAGCTTCCACTTCTC 805
QY 204 AsnTrpAlaGlyLeuAlaArgSerIleTrpLeuTySerValProGlnGlnHisIleGln 223
Db 806 AACTATGCGGGGTGACCGGCTGTGTGCTGTACACCACTTACCACTTATATCAT 865
QY 224 AspGluThrValValThrAspValAspGlyAspAsnGlyLeuIleAsnTrpGluValGlu 243
Db 1644 ACAGCTCTCTCGAGAAATTATCATTTTGTATCTCGATGAGAAAGAAAGATATGTATC 1903

```



QY 585 G1G1uH1sVal1TTPaenPheal1AaspPheG1nThraenLeuG1Y1le1leAryVal1Aasp 604  
 DB 1904 GGAAGCTCTACTGAGATTTTCTGACTTATGACGAAACCATGACATGAGAGTAC 1963  
 QY 605 G1YAsn1yLysG1YVal1PheThraArgAspArg1YsProlYsAla1Aa1Ahi1SerLeu 624  
 DB 1964 GGAACAGAAAGGAGATCTTCACTCGACAGAAACCACCAAGATGCGAGCTTATTTTG 2023  
 QY 625 ArgAlaArg1TPrh1Ser1le 631  
 DB 2024 CGAGAGATATCTGAGAGATT 2044  
 RESULT 2  
 CD014094  
 LOCUS CD014094 2095 bp mRNA linear EST 21-OCT-2003  
 DEFINITION 90135027 Single gene library Homo sapiens cDNA, mRNA sequence.  
 ACCESSION CD014094  
 VERSION CD014094.1 GI:37777623  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 BUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 2095)  
 JIN, P., FU, G. K., WILSON, A. D., YANG, J., CHIEN, D., HAWKINS, P. R.,  
 AU-YOUNG, J. and STUVE, L. L.,  
 PCR isolation and cloning of novel splice variant mRNAs from known  
 drug target genes  
 Genomics 83 (4), 566-571 (2004)  
 TITLE JOURNAL  
 COMMENT Contact: Jin, P.  
 Incyte Corporation  
 3160 Porter Drive, Palo Alto, CA 94304, USA  
 Tel: 650 621 8639  
 Fax: 650 621 8965  
 Email: pjin@incyte.com.  
 location/Qualifiers  
 1. 2095  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Single gene library"  
 /note="Vector: pDrive Cloning Vector; RT-PCR was performed  
 using gene-specific primers flanking the open-reading  
 frame. PCR products were subcloned into pDrive Cloning  
 Vector and sequenced completely using M13 forward and  
 reverse primers. Sequencing gaps were closed by  
 re-sequencing using primers flanking the gapped areas."  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 4,21e-129 Length: 2095  
 Score: 1253.50 Matches: 266  
 Percent Similarity: 57.51% Conservative: 90  
 Best Local Similarity: 42.97% Mismatches: 187  
 Query Match: 37.37% Indels: 77  
 Gaps: 13  
 US-10-757-093-4 (1-634) x CD014094 (1-2095)  
 QY 42 ProG1nA1rTh1Ser1Ser1ArgG1uLeuVal1Aa1nLeuAaP1YLeuT1rP1yS1PheAl1eU 61  
 DB 101 CCCGAGAGAGCCCGCTGCGGAGCTGCAAGAGCTGACCGCTCTTGAGCTTC----- 154  
 QY 62 A1s1erG1YLeuAa1rTh1r-----A1aG1nProT1rTh1A1Pro 75  
 DB 155 CGGCGCCACTTCTCTGACACCGACGCGGGGCTTCGAGGAGCGGTGACCGGCGCGCG 214  
 QY 76 LeuP1roLYs1eLY-----LeuG1uCyA1rProVal1ProAl1aSer1Y1rAa1rP1ePhe 92  
 DB 215 CTGTGGAGTCAAGCCCGCCACCGTGCATGCTCCAGTCTTCCACCTCAATGACATAGC 274  
 QY 93 I1e1SerArg1u1Le1H1sA1rP1H1sVal1G1Y1rVal1Y1rY1rG1nA1rG1uAl1eVal 112

DB 275 CAGAGCTGGCTGTGGGCGCATTTTGTGGCTGGGTGGTGAACGAGGAGGTATCTG 334  
 QY 113 ProLYsG1Y1rP1erG1nG1u-----Arg1Y1rLeuVal1rG1aG1uSer1A1rTh 129  
 DB 335 CCGAGCGAGAGTGAACCCAGGACCTGCCGACACAGAGTGTGTGAGATTGGCAATGCCAT 394  
 QY 130 H1eH1eG1YArg1Le1YrVal1Aa1nA1rG1uVal1A1aG1uH1sVal1G1Y1Y1rTh 149  
 DB 395 TCCTATGCCATGCTGTGGGTGAATGGGCTGCAGACCGCTAGAGCATGAGGGGCTACCTC 454  
 QY 150 ProPheG1uA1aA1rP1al1rTh1rG1uVal1A1aProG1Y-----G1uY1sPheArg 166  
 DB 455 CCCTTGAGGCGGACATACAGCAACCTGTGTCAAGTGGGGCCCTGCCCTCCGCTCCGA 514  
 QY 167 LeuTh1r1leG1YVal1e1nA1nG1uLeuTh1rH1eG1uTh1r1leProP1roG1Y1s1le--- 185  
 DB 515 ATCATATCCGCTCCACCAACACATCATACCCGCCACACCTGCGACAGGAGCATTCGA 574  
 QY 186 -----Th1rTh1rG1Aa1a1rTh1rG1Y1YsArg1leG1nTh1rY1rG1nH1sA1rP 201  
 DB 575 TACCTGACTGACCTCCAAATATCCCAAGGTTACTTGTCCAGAACATATTTTGAC 634  
 QY 202 PheTYrAsn1YrA1aG1YLeuA1aA1rG1er1le1rP1eY1rX1erVal1ProG1nG1H1s 221  
 DB 635 TTTTCACTACGCTGAGCTGACGCGGTGTGTCTGTGTAACAGACACCATCCACTAC 694  
 QY 222 I1eG1nA1rP1le1Th1rVal1Val1Th1rA1rP1a1rP1a1rG1Y1A1rP1a1n1Y1rG1u 241  
 DB 695 ATGATGATCATCACCGCTCACACACACGAGTGGGCTGGGCTGGGAAATTACAG 754  
 QY 242 ValG1uVal1A1aA1nG1nTh1rTh1rG1nG1nG1n1e1SerVal1leA1rP1a1rG1uA1rP1 261  
 DB 755 ATCTGTGTCAAGGCGAGTAACTGTTCATAGTGAAGTGGCTTTTGATGCAAGAAAC 814  
 QY 262 A1a1leVal1A1a1YsA1a1SerG1YAlaG1nG1YTh1rVal1Th1r1lePro1erVal1Y1sLeu 281  
 DB 815 AAGGTGTGCGAATGAGACTGGAGCCAGGCGCAACTTAAGTGCAGAGTGCAGCTC 874  
 QY 282 T1rG1nPro-----G1YAla1Y1rLeu1Y1rG1nG1nVal1A1n1le 296  
 DB 875 TGGTGGCTGTACCAAGAGCCAGTCTCTCATATGATGGAAGACCTTTCTATTTCCAGGT 934  
 QY 297 ValG1Y-----Ser1erG1YAs1rVal1A1rP1Th1rY1rA1nLeuA1a1rTh1rG1YAla1rG 314  
 DB 935 ACTGCACAGACTGACCTGGGCTGTGTGACTTCTTACACACTCCCTGTGGGAAATCCG 994  
 QY 315 Th1rVal1YsVal1A1aG1YSerG1nPh1eLeu1leA1nG1Y1YsA1rProP1eY1rP1eTh1rG1Y 334  
 DB 995 ACTGTGGCTGTACCAAGAGCCAGTCTCTCATATGATGGAAGACCTTTCTATTTCCAGGT 1054  
 QY 335 PheG1Y1YH1eG1uA1rP1Th1rA1a1rG1Y1YsG1YH1eA1rProA1a1rY1rMetVal 354  
 DB 1055 GTCAAAAGATGAGATGCGACATCCGAGGAAAGGCTTCCATCGCGCTGTGGTG 1114  
 QY 355 H1eA1rP1eG1nLeuMet1YrP1leG1YAla1A1nSer1Phe1rG1H1s1Y1rPro 374  
 DB 1115 AAGGACTTCAACTGCTGCTGGCTGTGGGCCAAGCGCTTCCGTACAGCCACTACCCC 1174  
 QY 375 TYrA1aG1uG1uVal1Me1A1rP1eA1aA1rP1A1rG1Y1leVal1Val1leA1rP1uTh1r 394  
 DB 1175 TATGAGAGAAATGATGCAATGATGTGACCCGTATGAGATTTGTGTCATGATGATGT 1234  
 QY 395 ProAlaVal1G1YLeuA1n1leA1a1eUeUeG1YAla1Ser1uSer1uG1YAla1ProG1nTh1r 414  
 DB 1235 CCGGCGGTGGCTGGCGCTG-----CCGAGTTTC 1264  
 QY 415 PheTh1rProA1rA1a1leA1nA1rP1Y1rTh1nG1nG1uA1H1e1YsG1nA1a1leA1rG1u 434  
 DB 1265 TTC-----AACCAAGTTTCTGTGCATCACCACTGCGAGGTGAGTGAAGAA 1309  
 QY 435 Leu1leA1a1rG1A1rP1YsA1nH1sA1rVal1A1Me1T1rP1er1leA1a1nG1uPro 454

Db	Accession	Source	Organism	Reference	Authors	Title	Journal	Comment
Db	1310	GTGGTGGATGAGGACAAAGAACCCCGGGGCTGTGATGTGGTCTGGCCAAAGACCT	1365					
Qy	4455	AlaSerHisGluAspGlyAlaArgGlyTyrPheGluProLeuThrAsnLeuThrArgGln	474					
Db	1370	GGCTCCACCTAGAAACTGTCTGGCTACTTCTT-GAA-----	1400					
Qy	4475	LeuAspProThrArgProIleThrPheAlaAsnValGlyThrAlaThrTyrGlnLeuAsp	494					
Db	1405	-----GCTCCGAT-----	1413					
Qy	4495	ArgLleSerAspLeuPheAspValSerCysIleAsnArgTyrPheGlyTyrTyrSerGln	514					
Db	1414	-----GTGATGTGATCTGTTTGAACAGCTACTACTCTGGATATACAGAC	1455					
Qy	515	ThrGlyAspLeuGlnGlnAlaGlnAlaLeuGlnIleuArgLeuHisGlyTyrGlnIleu	534					
Db	1459	TACGGGCACTTGGAGTGTGATTCAGCTGCAGCTGGCCACCATTGGAGAAGCTGATTAAG	1518					
Qy	535	LysPheHisArgProIleValMetThrGlyTyrGlyAlaAspThrLeuAlaGlyLeuHis	554					
Db	1519	AAGTATCAGAAAGCCCATTTATTCAGAGCGAGTATGAGCAGAAACGATTCACAGGTTTAC	1578					
Qy	555	SerLleLeuGlyLeuProTyrSerGlnGluPheGlnValGlnMetLeuAspMetTyrHis	574					
Db	1579	CAGGATCCACTCTGATGTTCTCACTAAGATACAGAAAGCTGCTGACAGACATACCAT	1638					
Qy	575	ArgValIleAsp-----ArgLleGluSerMetIleAlaGlyIleHisValIleThrAsnPheAla	592					
Db	1639	CTGGGCTCGGATCAAAAACGAGAAATAACGTCGTGGAGAGCTCATTTGGAAATTTGGCC	1698					
Qy	593	AspPheGlnThrAsnLeuGlyTyrLleIleArgValAspGlyAsnLysGlyValPheThr	612					
Db	1699	GATTTTCATGACTGAAACAGCTCACCGACGAGAGTGTCTGGGAAATTAATAAGGGATCTTCACT	1758					
Qy	613	ArgAspArgLysProLysAlaAlaAlaHisSerLeuArgAlaArgTyrThrSerLle	631					
Db	1759	CGCGAGAGACAAACCAAAAGTGCAGCGCTTCTTTTGCAGAGAGATCTGAGAAATTT	1815					
RESULT 3								
LOCUS	CD014092	1995 bp	mRNA	linear	EST 21-OCT-2003			
DEFINITION	90134967	Single gene library	Homo sapiens	cdna	mRNA sequence.			
ACCESSION	CD014092							
VERSION	CD014092.1	GI:37777621						
KEYWORDS	EST.							
SOURCE	Homo sapiens (human)							
ORGANISM	Homo sapiens							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.							
AUTHORS	1 (bases 1 to 1995)							
TITLE	Jin,P., Fu,G.K., Wilson,A.D., Yang,J., Chien,D., Hawkins,P.R., Au-Young,J. and Stuve,L.L.							
JOURNAL	PCR isolation and cloning of novel splice variant mRNAs from known drug target genes							
COMMENT	Genomics 83 (4), 566-571 (2004)							
	Contact: Jin, P.							
	Incyte Corporation							
	3160 Porter Drive, Palo Alto, CA 94304, USA							
	Tel: 650 621 8639							
	Fax: 650 621 8965							
	Email: pjmin@incyte.com							
FEATURES	Location/Qualifiers							
source	1..1995							
	/organism="Homo sapiens"							
	/mol_type="mRNA"							
	/db_xref="taxon:9606"							
	/clone_lib="Single gene library"							
	/note="Vector: pDrive Cloning Vector; RT-PCR was performed using gene-specific primers flanking the open-reading frame. PCR products were subcloned into pDrive Cloning Vector and sequenced completely using M13 forward and reverse primers. Sequencing gaps were closed by re-sequencing using primers flanking the gapped areas."							

Alignment Scores:	2.62e-120	Length:	1995
Pred. No.:	1175.00	Matches:	251
Score:	55.92%	Conservative:	89
Percent Similarity:	41.28%	Mismatches:	181
Best Local Similarity:	35.03%	Indels:	89
Query Match:	6	Gaps:	12

US-10-757-093-4 (1-634) x CD014092 (1-1995)

QY	42	ProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu	61
DB	101	CCCCAGAGAGACCCCGTCGGCGAGTCGACAGAGACTGACGCGCTCTGGACGCTTC-----	15
QY	62	AlaSerGlyLeuAsnAspThr-----AlaGlnProTrpThrAlaPro	75
DB	155	CGCGCGCATCTCTGCAACACGACCGCGGCGCTTCGAGAGACGTCGACCGGCGCG	21
QY	76	LeuProLysGly-----LeuGlnCysProValProAlaSerTrpAsnAspIlePhe	92
DB	215	CTGCGGAGTCAGGCCCCACCGTCGACATGCCAGTCCTCCAGCTTCACATGACATCAGC	27
QY	93	IleSerArgGluIleHisAspHisValGlyTrpValTrpTrpGlnArgGluValIleVal	11
DB	275	CAGAGACTGCGCTGCGCGCATTTTTCGCGGTGGTCGTGTGTCGACACGGAGGTGATCTCG	33
QY	113	ProLysGlyTrpSerGlnGluArgTrpLeuValArgAlaGlnSerAlaThrHisGly	13
DB	335	CCGAGCGCATGACCCACGAGC-----	35
QY	133	ArgIleTrpValAsnAsnArgLeuValAlaGlnHisValGlyGlyTrpThrProPheGlu	15
DB	356	-----CTGCGGCACAGAGTGGTG-----	37
QY	153	AlaAspValThrGlnLeuValAlaProGlyGlyLysPheAspGlyThrIleGlyValAsn	17
DB	374	-----CTGAGGATTTGGCAGTGGCC	39
QY	173	AsnGluLeuThrHisGlnThrIleProProGlyLysIleThrGlyAsnAlaThrGly	19
DB	392	CATTC-CTATGCGATCGTGATCCAG-GGTACTTT-----	42
QY	193	LysArgGlnGlnThrTrpGlnHisAspPheTrpAsnTrpAlaGlyLeuAlaArgSerIle	21
DB	429	-----GTCCAGAACACATATTTTCACTTTTTCACATCGCTGACGTGACGCGTCTGTA	48
QY	213	TrpLeuTrpSerValProGlnGlnHisIleGlnAspIleThrValIleThrAspValAsp	23
DB	483	CTTCTGTGACACGACCCACACCACTACATCATGACATCACCGTCGACACGAGCTGAG	54
QY	233	GlyAspAsnGlyLeuIleAsnTrpGluValGluValAlaAsnGlnThrTrpGlyGlnIle	25
DB	543	CAGAGACAGTGGGTGTGTGATTAACAGATCTCTGCAAGGCGCAGTAACGTTCACAGTTG	60
QY	253	GlnIleSerValIleAspGlnAspGlyAlaIleValAlaIleValAsnAspGlyAlaGlnGly	27
DB	603	GAGTGCCTTTTGGATGTCAGAAAAACAAGTCGTGGGAGATGGAGTGGACCCAGGCG	66
QY	273	ThrValThrIleProSerValLysLeuTrpGlnPro-----GlyAlaAla	28
DB	663	CAACTTAAGTCCCGAGTGTGACGCTCTGTGGCGCTGATGACGACAGCCCGCTGCC	72
QY	288	TrpLeuTrpGlnLeuGlnValAsnIleValGly-----SerSerGlyAspValValAsp	30
DB	723	TATCTGTATTCATGAGGTGACGTGACTGACACAGACGTCACGTGGGCGCTGTGTGAC	78
QY	306	ThrTrpAsnLeuAlaThrGlyValArgTrpValLysValAlaGlnSerGlnPheLeuIle	32
DB	783	TCTTACACACTCCCTGTGGGATCGGACACTGTGGCTCTCACCAAGACCGATTCTCATC	84
QY	326	AsnGlyLysProPheTrpThrGlyPheGlyLysHisGlnAspThrAlaValArgGly	34

[illegible]

ORGANISM	Mus musculus Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1
AUTHORS	Carninci, P. and Hayashizaki, Y.
JOURNAL	High-efficiency full-length cDNA cloning
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)
PUBMED	99279253
REFERENCE	10349636
AUTHORS	2
TITLE	Garcinzi, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, J., Nishi, K., Kitunai, T., Taeshiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Niimine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kasihagi, K., Fujiwara, K., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Osawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
TITLE	Riken integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	4
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
MEDLINE	5
PUBMED	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 5 (bases 1 to 2274)
REFERENCE	Adachi, T., Aizawa, K., Akinura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imocant, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Koijima, Y., Kondo, S., Komno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ono, M., Ohsato, N., Okezaki, Y., Satoh, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Saeki, D., Shibata, K., Shingawa, A., Shiraki, T., Sojabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasuniishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suicho-cho, Tsukuba-shi, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in RIKEN Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/ Location/Qualifiers 1..2274 /organism="Mus musculus"
FEATURES	
SOURCE	

```

/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM,DB:A530072005"
/db_xref="taxon:10090"
/clone="A530072005"
/sex="male"
/tissue_type="aorta and vein"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
14..11778
/note="beta-glucuronidase structural (MED|MG1:95874,
GB|NM_010368, evidence: BLASTN, 99%, match=733)
putative"
polya_signal
2260..2265
/note="putative"
polya_site
2274
/note="putative"

ORIGIN

Alignment Scores:
Pred. No.: 1,73e-119 Length: 2274
Score: 1168.50 Matches: 249
Percent Similarity: 55.76% Conservative: 90
Best Local Similarity: 40.95% Mismatches: 179
Query Match: 34.84% Indels: 90
DB: 3 Gaps: 13

US-10-757-093-4 (1-634) x AK041058 (1-2274)

QY 42 ProGlnRgThrSerSerArgGluLeuValAsnLeuAspGlyLeuThrPlySphenAlaLeu 61
D 101 CCGAAGAGAGAGCCCGTGGCGGAGACTCAAGCGCTGAGACGACTGTGGCACTTCGGCGCC 160
QY 62 -----AlaSerGlyLeuAsnAspThrAlaGlnProTrp---ThrAlaProLeuPro 77
D 161 GACCTCTGCAACACCGGCTGCAGGCTTTCGAGAGCAATGTCATCCGGACCGCTTACGG 220
QY 78 LysGly-----LeuGluCysProValProAlaSerTyrsnAspIlePheIleSer 94
D 221 GAGTCGGGCCCAAGCTTGTGACATGCTGTCTCTTCTTCAATGATCATCACCACCAAGA 280
QY 95 ArgGluIleHisAspHisValGlyTrpValTyrtGlnArgGluValIleValProLys 114
D 281 GCAGCCCTTCGGACCTTATTTGGCTGGTGGTGGATGACGGAGCAATCCGCGACCG 340
QY 115 GlyTrpSerGlnGlu-----ArgTyrlLeuValArgAlaGluSerAlaThrHis 131
D 341 CGATGGACCCCAAGATACCGACATGAGAGTGCTGTGAGATCAACAGTGCC---CATTTAT 397
QY 132 GlyArgIleTyrlAsnAsnArgLeuValAlaGluHisValGlyGlyTyrlThrProPhe 151
D 398 TATGCACTTGTGATTCGCA-----
QY 152 GluAlaAspValThrGluLeuValAlaProGlyGlyCys-PheArgLeuThrIleGlyVal 171
D 419 -----GGTTACTTGTTCGACGACACAAAGCTTT-----
QY 171 LaenAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThrGlyAsnAlaThr 191
D 446 -----
QY 191 rGlyLysArgIleGlnThrTyrlGlnHisAspPheTyrlAsnTyrlAlaGlyLeuAlaArgSe 211
D 447 -----GACTTCTTCAACTATGCGGAGCTGCATCGATC 478
QY 211 rIleTrpLeuTyrlSerValProGlnGlnHisIleGlnAspIleThrValValThrAspVal 231
D 479 TGGGTCCTCTATACACCCCTTACCATCATCATGATATCATCTGTGATCATCAATGT 538
QY 231 LaerGlyAspAsnGlyLeuIleAsnTyrlGluValGluValAlaAsnGlnThrThrGlyGly 251
D 539 GGAAGCAAGCATCGGCTGTGTGACCTATTCGATTCGTGACGGGACATGAACTTTCCA 598

```

```

QY 251 nIleGlnIleSerValIleAspGluAspGlyAlaIleValAlaIleValAlaSerGlyValaGly 271
D 599 GCTAAGAGTGCAACTTTGGATGAGGGTGCAAAAGTCGTGGCCATGAGGACAGGAGCAACA 658
QY 271 nGlyThrValThrIleProSerValLysLeuThrPro-----GlyVal 286
D 659 GGGTCAACTTCAGGTTCCAGTCCAGTCCCACTCTGTGGCTTACCTGATCATGACATGCC 718
QY 286 aAlaTyrlLeuTyrlGlnLeuGlnValAsnIleValGlySerSerGlyAspValValAspThr 306
D 719 AGCCTACATGATCTCTTGAGAGGTGAAGTACA---ACAACGTAGCTGTGATGACTCA 775
QY 306 rTyrlAsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIleAs 326
D 776 CTACACCTTCCTATCGGAGATTGCAACAGTGGCTGTCAACAAAGACAAAGTCTTCATAA 835
QY 326 nGlyLysProPheTyrlPheThrGlyPheGlyLysHisIleGluAspThrAlaValArgGly 346
D 836 CCGGAAGCCCTCTATTTCCAAAGGCTCAATAGCAAGAGATTGATATCCAGAGGAA 895
QY 346 rGlyHisAspProAlaTyrlMetValHisAspPheGlnLeuMetTyrlTrpIleGlyValAs 366
D 896 AGGCTTCGACTGCGCGCTGTGTAAAGGATTTCAACCTGCTCCGTGGCTCGGGCAAA 955
QY 366 nSerPheArgTrpSerHisTyrlProTyrlAlaGluGluValMetAspPheAlaAspArgAs 386
D 956 TTCCTTTCTTACCAAGCCTATCTCCCTACTCAGAGGAGGACTTACGCTCTGTGACCGATA 1015
QY 386 nGlyIleValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeuMetGlyVal 406
D 1016 CCGGATTTGGTATCATGATGATGATGCCGCTGGGCAATTGGCTA----- 1061
QY 406 lSerGluSerGlyAlaProGlnThrPheThrProAspAlaIleAsnAspLysThrGlnGly 426
D 1062 -----CTCAGAGTTTGGCAACCAAGCACTTCCGCGAC----- 1094
QY 426 uAlaHisGlyAlaIleArgGluLeuIleAlaArgAspLysAsnHisAlaSerValIle 446
D 1095 -----CACCTAGAGGATGAGAGAGAGCTGTTCGCGGACAAATAACCTCGCGGTTGT 1150
QY 446 lMetTrpSerIleAlaAsnGluProAlaSerHisGluAspGlyValArgGlyTyrlPheGly 466
D 1151 GATGGTGTCTGGGCAATGAGCCCTTCTGCTGTCAAAACCCCGCATATTACTTTAA 1210
QY 466 uProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnVal 486
D 1211 GACGCTGATCACCCACACCAAGCCCTGAGCTCAGCCGCTCCGTGACCTT-----GT 1264
QY 486 lGlyThrAlaThrTyrlGlnLeuAspArgIleSerAspLeuPheAspValSerCysIleAs 506
D 1265 GAGCAACGCCAAATATATGATGACAGCTGGGGGCCCCGTAAGTGTATCTCTGTAA 1324
QY 506 nArgTyrlPheGlyTyrlTrpTyrlSerGlnThrGlyAspLeuGluGluAlaIleAlaLeuGly 526
D 1325 CAGCTACTTTCTTGGATATCATGACTATGGGCATTTGAGGTGATTCAGCCACAGCTGAA 1384
QY 526 uLysGluLeuHisGlyTyrlGlnGlnLysPheHisArgProIleValMetThrGlyTyrlGly 546
D 1385 TAGCCAGTTTGAAGAACTGTATTAAGACCATCAGAACCCGATTTTCCAGACGAGATGG 1444
QY 546 yAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGlnGluPheGly 566
D 1445 AGCAGACGCAATCCAGAGGATCCACAGAGACCCGCTCGCATGTTCAGTGAAGGTATCCA 1504
QY 566 nValGlnMetLeuAspMetTyrlHisArgValPheAsp-----ArgIleGluSerMetAl 584
D 1505 GAAGCTGTGTGGAGAAATTACATTCATTCGATTCAGAAACGTAAATAAATACGTGT 1564
QY 584 aGlyGlnHisValTyrlAsnPheAlaAspPheGlnThrAsnLeuGlyTyrlIleArgValAs 604
D 1565 CCGAGAGCTATATCGAAATTTTCGCGCACTTCAAGACAAACAGTCAACCGCTGAGAGTAT 1624
QY 604 pGlyAsnLysLysGlyValPheThrArgAspArgLysProLysAlaAlaHisSerIle 624

```

```

Db      1625 CGGAACAGAGAGGATCTTCACTGCGCCAGAGACGCCAAACTCGGCTTATTTT 1684
Qy      624  |||||
        |||||
        |||||
Db      1685 GCGAGAGAGATCTGAGAGATT 1706

RESULT 5
CD014093      1853 bp      mRNA      linear      EST 21-OCT-2003
LOCUS      90135266 Single gene library Homo sapiens cDNA, mRNA sequence.
DEFINITION      CD014093
ACCESSION      CD014093
VERSION      CD014093.1 GI:37777622
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1853)
Jin, P., Fu, G.K., Wilson, A.D., Yang, J., Chien, D., Hawkins, P.R.,
Au-Yang, J. and Stuve, L.L.
PCR isolation and cloning of novel splice variant mRNAs from known
drug target genes
Genomics 83 (4), 566-571 (2004)
Contact: Jin, P.
Incyte Corporation
3160 Porter Drive, Palo Alto, CA 94304, USA
Tel: 650 621 8639
Fax: 650 621 8965
Email: pj@incyte.com.
Location/Qualifiers
1..1853
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="Single gene library"
/notes="Vector: pDrive Cloning Vector; RT-PCR was performed
using gene-specific primers flanking the open-reading
frame. PCR products were subcloned into pDrive Cloning
Vector and sequenced completely using M13 forward and
reverse primers. Sequencing gaps were closed by
re-sequencing using primers flanking the gapped areas."

ORIGIN

Alignment Scores:
Pred. No.:      8.53e-106      Length:      1853
Score:      1045.50      Matches:      227
Percent Similarity:      50.90%      Conservative:      84
Best Local Similarity:      37.15%      Mismatches:      159
Query Match:      31.17%      Indels:      141
DB:      6      Gaps:      13

US-10-757-093-4 (1-634) x CD014093 (1-1853)
Qy      42  ProGlnArgThrSerSerArgGluLeuValAsnLeuAspGlyLeuTrpLysPheAlaLeu 61
        |||||
        |||||
        |||||
Db      101  CCCAGAGAGAGCCCGTGGCGGAGTGCAGAGAGCTGAGCGCCCTGAGACTTC----- 154
        |||||
        |||||
        |||||
Qy      62  AlaSerGlyLeuAsnAspThr-----AlaGlnProTrpThrAlaPro 75
        |||||
        |||||
        |||||
Db      155  CGGCGCGACTTCTCTGACCAACCGCGCGGCGCTTCGAGAGCGAGTGTAACCGCGCGCG 214
        |||||
        |||||
        |||||
Qy      76  LeuProLysGly-----LeuGluCysProValProAlaSerTyrAsnAspIlePhe 92
        |||||
        |||||
        |||||
Db      215  CTGTGGAGTCAAGGCCCGCCACCGTGCAGATGCAGATTCCTCCAGCTTCAATGACATAGC 274
        |||||
        |||||
        |||||
Qy      93  ILeSerArgGluIleHisAspHisValGlyTrpValTyrTrpGlnArgGluValIleVal 112
        |||||
        |||||
        |||||
Db      275  CAGGACTGGCGCTCTGCGCGCATTTTGTGCGCTGGGTGGTGCAGCAACGAGAGATCTCTG 334
        |||||
        |||||
        |||||
Qy      113  ProLysGlyTrpSerGlnGlu-----ArgTyrLeuValArgAlaGlnSerAlaThr 129
        |||||
        |||||
        |||||
Db      335  CCGGAGCGATGAGACCCAGAGACTGCGCGCAACAAGAGTGTGCTGAGAGATTGCGACATGCC-- 391

```

```

Qy      130  HLeIeGlyArgIleTyrValAsnAsnArgLeuValAlaGluHisValGlyGlyTyrThr 149
Db      391  -----
Qy      150  ProPheGluAlaAspValThrGluLeuValAlaProGlyGlyLysPheArgLeuThrIle 169
Db      391  -----
Qy      170  GlyValAsnAsnGluLeuThrHisGluThrIleProProGlyLysIleThrThrGlyAsn 189
Db      391  -----
Qy      190  AlaThrGlyLysArgIleGlnThrTyrGlnHisAspPheTyrAsnTyrAlaGlyLeuAla 209
Db      392  -----CATTCCTATGCC-----403
Qy      210  ArgSerIleTrpLeuTyrSerValProGlnGlnHisIleGlnAspIleThrValAlaThr 229
Db      404  -----ATCGTT-----409
Qy      230  AspValAspGlyAspAsnGlyLeuIleAsnTyrGluValGluValAlaAsnGlnThrThr 249
Db      410  -----GGCGTGGTGAATTACAGATCTCTGCAAGGCGCAGTAACCTG 451
Qy      250  GlyGlnIleGlnIleSerValIleAspGluAspGlyValIleValAlaLysAlaSerGly 269
Db      452  TTCAGTGGAGATGGCTCTTTGGATGCGAAGAAACAAAGCTGGCGAATGGACATGGG 511
Qy      270  AlaGlnGlyThrValThrIleProSerValLysLeuTrpGlnPro-----284
Db      512  ACCGAGGGCAACTTAAGGCGCAGGTCAGAGCTCTGGTGGCCGTATCGATGACAGAA 571
Qy      285  GlyAlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValGly-----SerSerGlyAsp 302
Db      572  CGCGCTGCTATCTGATTCATTTGAGTGAAGGTCAGCTACCTGACAGACGTCACCTGGGCT 631
Qy      303  ValValAspThrTyrAsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGln 322
Db      632  GTTCTGACTTCTTACACACTCTCTGGGGATGCGACCTGTGCTGACCAAGAGCGAG 691
Qy      323  PheLeuIleAsnGlyLysProPheTyrPheThrGlyPheGlyLysHisGluAspThrAla 342
Db      692  TTCTCATCATGATGGAAACCTTTCTATTTCCAGGGTCAACAAGATGATGAGTGGGAC 751
Qy      343  ValArgGlyLysGlyHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTrp 362
Db      752  ATCCGAGGGAAGGCGCTTCGACTGCGCGCTGCTGTGTAAGGACTTCAACCTGCTCGTGG 811
Qy      363  IleGlyValAsnSerPheArgThrSerHisTyrProTyrAlaGluGluValMetAspPhe 382
Db      812  CTGGTGCCAAACGCTTCGTTACGACGCTGCGCTGCTGTGTAAGGACTTCAACCTGCTCGTGG 871
Qy      383  AlaAspArgAsnGlyIleValValIleAspGluThrProAlaValGlyLeuAsnIleAla 402
Db      872  TGTGACCGCTATGGAGTGTGTGTCATGATGATGTCGCGCGCTGGCGCTGGCGCTG 928
Qy      403  LeuMetGlyValSerGlnSerGlyAlaProGlnThrPheThrProAspAlaIleAsnAsp 422
Db      929  -----CCGCACTTCTTC-----AACAC 946
Qy      423  LysThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsnHis 442
Db      947  GTTCTCTGCATACCAACATGACGAGTGAAGAGAGTGGTGGTGAAGCAAGAACAC 1006
Qy      443  AlaSerValValMetTrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAlaArg 462
Db      1007  CCGCGGTGTATATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1066
Qy      463  GluTyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIleThr 482
Db      1067  TACTACTTGAAGATGTGTATGCTGTCAACCAATTCCTTGAACCCCTCCGCGCTGTGACC 1126

```

```

Qy 483 PheAlaAsnValGlyThrAlaThrTyrglnLeuAspArgLleSerAspLeuPheAspVal 502
Db 1127 TTT-----GTGAGCACTCTACTATGAGCAGACAAAGGGGCTCCCTATGTGATGTG 1180
Qy 503 SerGylleAsnArgTyrglnTyrglnTyrglnTyrglnTyrglnTyrglnTyrglnTyrgln 522
Db 1181 ATCTGTTGAAACAGTACTACTCTGATATCAGCACTACGGGCACTGGAGTTGATTCAG 1240
Qy 523 AlaAlaLeuGlnGlnLeuHisGlyTyrglnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 542
Db 1241 CTGCGAGCTGGCCACCACTTGGAGAACTGGTATTAAGATTCAGAACCCATTATTCAG 1300
Qy 543 ThrGlnTyrglnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 562
Db 1301 AGCGAGTATGGAGCAGAAACGATTCAGGGTTTCCAGAGATTCACCTCTATGTTCACT 1360
Qy 563 GlnGlnPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 580
Db 1361 GAAGAGTACCAAGAAAGTCTGCTAGACAGTACCATCTGGGCTGATCAAAAACGAGA 1420
Qy 581 GlnSerMetAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 600
Db 1421 AAATCGTGGTGGAGAGACTCATTTGGATTTTGGCCATTTTCAGTCACTGAAACAGTACCG 1480
Qy 601 IleArgValAspGlyAsnGlyValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 620
Db 1481 ACGGAGTCTGGGGGATTAAGGGGATCTTCACTCGGAGAGACAAACCAAAAGTGC 1540
Qy 621 AlaHisSerLeuArgAlaArgTrpThrSerIle 631
Db 1541 GCGTTCCTTTGCGAGAGATCTCGAAGAT 1573

RESULT 6
CR593823 1377 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CSODL001Y21 of B cells (Ramos cell line)
DEFINITION Cc 25-normalized of Homo sapiens (human).
ACCESSION CR593823.1 GI:50474630
VERSION CR593823.1
KEYWORDS HTC; CNSLT cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1. (bases 1 to 1377)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Parade Avenue
2 (bases 1 to 1377)
Genoscope.
REFERENCE Direct Submission
AUTHORS Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source location/Qualifiers
1..1377
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODL001Y21"
/tissue_type="B cells (Ramos cell line) Cc 25-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
* Alignment Scores:

```

```

Pred. No.: 8,28e-95 Length: 1377
Score: 946.00 Matches: 186
Percent Similarity: 61.48% Conservative: 63
Best Local Similarity: 45.93% Mismatches: 130
Query Match: 28.21% Indels: 26
DB: 3 Gaps: 6

US-10-757-093-4 (1-634) x CR593823 (1-1377)
Qy 236 GlyLeuIleAsnTyrglnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 255
Db 1 GGGGTGGTGAATTAACGATCTCTGTCAGGGGAGAACCTGTCTCAAGTTGAGAGCGCT 60
Qy 256 ValIleAspGlnAspGlnAlaIleValAlaIleValAlaSerGlyAlaGlnGlnTyrglnThr 275
Db 61 CTTTGGATGCAAAAACAAAGTCTGGGGAATGGAGTGGGACCCAGGCGCACTTAAG 120
Qy 276 IleProSerValIleLeuTrpGlnPro-----GlyAlaAlaTyrglnGlnGlnGlnGlnGln 290
Db 121 GTGCCAGGTCTCAGCTCTGCTGGGCGCTACCTGATGCAGAACGCCCTGCTATCTGAT 180
Qy 291 GlnLeuGlnValAsnIleValGly-----SerSerGlyAspValValAspThrTyrgln 308
Db 181 TCATTGAGAGTGGAGCTGACTGACAGACAGTCACTGGGGCTGTGCTGATCTTACACA 240
Qy 309 LeuAlaThrGlnValArgThrValIleValAlaGlnSerGlnPheLeuIleAsnGlnGln 328
Db 241 CTCTCTGGGGAGTCCGACGTGTGCTGCTGACCAAGCGCACTTCTCATCAATGGGAA 300
Qy 329 ProPheTyrglnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 348
Db 301 CTTTCTATTTCACAGGTGTCAACAGCATGAGATGGCGATCTCGAGGAGAGGAGCTTC 360
Qy 349 AspProAlaTyrglnMetValHisAspPheGlnLeuMetIleTrpIleGlyAlaAsnSerPhe 368
Db 361 GACTGGCCGCTCTGCTGGTGAAGACTTCAACCTGCTTCCGCTGGTGGCAACGCTTTC 420
Qy 369 ArgThrSerHisTyrglnProTyrglnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 388
Db 421 CCGTACAGCCACTACCCCTATGCAAGAGAGAGAGTGCAGATGTGACCGCTATGGGAT 480
Qy 389 ValIleAlaAspGlnThrProAlaValGlnLeuAsnIleAlaLeuMetGlyValSerGln 408
Db 481 GGTGTCATGATGAGTGTGCCGGCGTGGCGCTGGCGCTG----- 519
Qy 409 SerGlnAlaProGlnThrPheThrProAspAlaIleAsnAspIleThrGlnGlnAlaHis 428
Db 520 -----CCGCACTTCTTC-----AACCAAGTTTCTTCATCACAC 555
Qy 429 IysGlnAlaIleArgGlnLeuIleAlaArgAspIleAsnHisAlaSerValIleMetTrp 448
Db 556 ATGCAAGTATGAGAAAGAGTGTGCTGAGGACAAAGCAACCCCGCTGTATGTGG 615
Qy 449 SerIleAlaAsnGlnProAlaSerHisGlnAspGlnValArgGlnTyrglnProLeu 468
Db 616 TCTGGGCAACGAGCCGCTGCCCTAGAACCTGCTGCTACTTGAATGATG 675
Qy 469 ThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThr 488
Db 676 ATCGTCAACCAAACTCTTGAACCCCTCCGAGCTGTGACTT-----GTAGCAAC 729
Qy 489 AlaThrTyrglnLeuAspArgLleSerAspLeuPheAspValSerGylleAsnArgTyrgln 508
Db 730 TCTAACTATGAGAGACAAAGGGGCTCCGATGTGATGTGATCTTTGAACAGTAC 789
Qy 509 PheGlnTyrglnTyrglnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 528
Db 790 TACTCTTGGTATCAGCACTACGGGCACTGGAGTTGATGACGTCGACGCCACCCAG 849
Qy 529 LeuHisGlnTyrglnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 546
Db 850 TTTGAGAACTGTATTAAGAGTATCAAGGCCCATTTATTCAGAGCGAGTATGAGAGCA 909

```

QY 549 ThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTrpSerGluGluPheGlnValGln 568  
 Db 910 ACATTTCGAGGCTTTCACCAAGATCCACTTATGTTCTTCAAGTACCAAGAAAGT 969  
 QY 569 MetLeuAspMetTyrHisArgValPheAsp-----ArgIleGluSerMetAlaGlyGlu 586  
 Db 970 CTCCTAGACAGATACATCTGCTGTGATCAAAAACGAGAAATATCTGTGTGGAGAG 1029  
 QY 587 HisValTyrPhePheAlaAspPheGlnThrAsnLeuGlyIleIleArgValAspGlyAsn 606  
 Db 1030 CTCATTGTGAATTTTGGCATTTTCATGACTGAACAGTACCGACGAGAGTGTGGGGAAT 1089  
 QY 607 LysValGlyValPheThrArgAspArgValProLysAlaIleAlaHisSerLeuArgAla 626  
 Db 1090 AAAAAGGAGATCTTCTACTCGGACAGACAAACAAAGTCCGCTTCTTTCAGAG 1149  
 QY 627 ArgTrpThrSerIle 631  
 Db 1150 AGATACCTGAGAGATT 1164  
 RESULT 7  
 CD503076 1124 bp mRNA linear EST 12-JUN-2003  
 LOCUS CDA60-C07.5, mRNA sequence.  
 DEFINITION CDA60-C07 5', mRNA sequence.  
 ACCESSION CD503076  
 VERSION CD503076.1 GI:31432977  
 KEYWORDS EST.  
 SOURCE Gasterosteus aculeatus (three spined stickleback)  
 ORGANISM Gasterosteus aculeatus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;  
 Gasterosteidae; Gasterosteus.  
 1 (bases 1 to 1124)  
 Kingley, D.M., Peichel, C., Balabhadra, S., Grimwood, J., Dickson, M.,  
 Schmutz, J., and Myers, R.M.  
 Expressed sequence tags from Gasterosteus aculeatus  
 Unpublished (2003)  
 Contact: Kingley, DM  
 HMMI and Department of Developmental Biology  
 Stanford University School of Medicine  
 Beckman Center B100, 279 Campus Drive, Stanford, CA 94305-5329, USA  
 Tel: 650 725 5954  
 Fax: 650 725 7739  
 Email: kingley@cmgm.stanford.edu  
 Plate: 60  
 High quality sequence stop: 782.  
 Location/Qualifiers  
 1..1124  
 /organism="Gasterosteus aculeatus"  
 /mol\_type="mRNA"  
 /strain="Salinas river, CA"  
 /db\_xref="taxon:69293"  
 /clone="CDA60-C07"  
 /sex="mixed male and female"  
 /tissue\_type="heads and internal organs combined"  
 /dev\_stage="adult"  
 /note="Vector: lambda ZAP Express/pBK-CMV; Site 1: EcoRI  
 (5' adaptor); Site 2: XhoI (3' linker primer); The mixed  
 organ cDNA library was generated using the ZAP-cDNA method  
 by Stratagene. First strand cDNA synthesis was primed with  
 a 50 bp linker primer containing an oligo dt sequence  
 preceded by a synthetic XhoI site. 5 prime adaptors were  
 used containing an EcoRI cohesive end. The finished cDNAs  
 were inserted in to the ZAP express vector  
 unidirectionally in the sense orientation with respect to  
 the lacZ promoter of pBK-CMV. An amplified library was  
 prepared from approximately 3 million primary clones in  
 the lambda ZAP Express vector. In vivo excision was then  
 used to generate individual pBK-CMV phagemid clones for  
 EST sequencing."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 7,26e-79 Length: 1124  
 Score: 803.00 Matches: 164  
 Percent Similarity: 60.748 Conservative: 48  
 Best Local Similarity: 46.994 Mismatches: 115  
 Query Match: 23.948 Indels: 23  
 DB: 6 Gaps: 5  
 US-10-757-093-4 (1-634) x CD503076 (1-1124)  
 QY 284 ProGlyAlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspVal 303  
 Db 8 CCAACCCAGGTTACTTCTTACTCTTTGGAGTTGTCTTACTGCTGCTAGAGATCTGCT 67  
 QY 304 Val--AspThrTyrAsnLeuAlaThrGlyValArgThrValLysValAlaGlySerGln 322  
 Db 68 TCCACGACGCTGTAACCTTACCAAGTCCGATCCGACGGTCCAGCGTTCACACACCAG 127  
 QY 323 PheLeuIleAsnGlyLeuProPheTyrPheThrGlyPheGlyLysIleGluAspThrAla 342  
 Db 128 TTCCTCATCAACAAAGCCCTTCTTCTCCACGAGTAATTAACACGAGACTCTGAT 187  
 QY 343 ValArgGlyLysGlyHisAspProAlaTyrMetValHisAspPheGlnLeuMetLysTrp 362  
 Db 188 ATTGAGGCAAGGCTGAGCTGCCCCCTCATGTGTAAGACTTTAATCTTATGAAGTGG 247  
 QY 363 IleGlyAlaAsnSerPheArgThrSerHisTyrProTyrAlaGluGluValMetAspPhe 382  
 Db 248 TTGGGGGCCCAACTGTTCCGACACCACTACCTTATGACAGAGATCTGACAGATG 307  
 QY 383 AlaAspArgAsnGlyIleValValIleAspGlnThrProAlaValGlyLeu--AsnIle 401  
 Db 308 TGTGACCGCATGTCATCGTGTGATGACAGATGCCCGGCTGGCATTAAGACATT 367  
 QY 402 AlaLeuMetGlyValSerGluSerGlyAlaProGlnThrPheThrProAspAlaIleAsn 421  
 Db 368 CGCAGTTTGA----- 379  
 QY 422 AspLysThrGlnGluAlaHisLysGlnAlaIleArgGluLeuIleAlaArgAspLysAsn 441  
 Db 380 AACGCTCCTTAACCATCATCATCGTGTGATGACGAGCTGTGATGCTGGGCAAGAAC 439  
 QY 442 HisAlaSerValValMetTrpSerIleAlaAsnGluProAlaSerHisGluAspGlyAla 461  
 Db 440 CATCCCTGTGTGTCATGTGTGATGAGCCCAAGACCCGGCTGCAGAGATGCTCTGCT 499  
 QY 462 ArgGluTyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIle 481  
 Db 500 GATTACTATTTCAAACCTTGATTAACATACCAAGAAATTGATCCAAACCCGCCGCTC 559  
 QY 482 ThrPheAlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAsp 501  
 Db 560 ACTTTT-----ATCACAGACAGTACTATGACAGGAGTAAGAGGCTCTACGTGAC 613  
 QY 502 ValSerCysIleAsnAspArgTyrPheGlyTyrPyrSerGlnThrGlyAspLeuGluAla 521  
 Db 614 GTCATCTGGCTAAACGTTACTTCTCTGTTACATGACCCCGGCCACCCGAGGTGATC 673  
 QY 522 GluAlaAlaLeuGluLysGluLeuHisGlyTyrGlnGluLysPheHisArgProIleVal 541  
 Db 674 CCATCCAGCTCAACACTGCTGTGAGAACTGTATCGAAGTAAGTAACGAAGCAACCATATC 723  
 QY 542 MetThrGluTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTrp 561  
 Db 734 CAGAGGGAATACGAGACGAGTGTGGTCCGGGCTTCACAGTGATCACCCGGATGTTT 793  
 QY 562 SerGluGluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAsp-----Arg 579  
 Db 794 ACTGAGAGATACCGAAGTGAATCTCGCAGAGCTACCAACCTGTTCGACCGAAAGG 853  
 QY 580 IleGluSerMetAlaGlyGluHisValTyrAsnPheAlaAspPheGlnThrAsnLeuGly 599

Db 854 AACGACATGCTATCGGCAACTCTGAACTTTGACACTTCTGACCCACAGGG 913  
 QY 600 ILeIleArgValAspGlyAsnLysGlyValPheThrArgAspArgLysProLysAla 619  
 Db 914 ATATGCGCGTGTGGGGAACAAAGGCTGCTCTGACGAGCA -AGGAGCCCAAGCAG 972  
 QY 620 AlaAlaHisSerLeuArgAlaArgTTP 628  
 Db 973 CAGCATCTCTGAAGGAGAGTACTGG 999

RESULT 8  
 BX363460  
 LOCUS  
 DEFINITION  
 Homo sapiens cDNA clone CS0DL001YM21 5-PRIME, mRNA sequence.  
 ACCESSION  
 BX363460.2 GI:46291531  
 VERSION  
 EST.  
 KEYWORDS  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 1051)  
 L1,W.B., Gruber,C., Jessee,J. and Polyes,D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 On May 5, 2003 this sequence version replaced gi:30372720.  
 COMMENT  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 4169.r  
 For more information about this cluster, see  
 http://www.genoscope.cns.fr/cdna?b=CS0DL001Y21&c=4169.r.  
 Location/Qualifiers  
 1..1051  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DL001YM21"  
 /cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /clone\_lib="RAMOS CELL LINE"  
 /cell\_line="RAMOS CELL LINE"  
 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES  
 source

ORIGIN

Alignment Scores:  
 Pred. No.: 2.72e-78 Length: 1051  
 Score: 797.50 Matches: 159  
 Percent Similarity: 59.50% Conservative: 57  
 Best Local Similarity: 43.80% Mismatches: 121  
 Query Match: 23.78% Indels: 26  
 DB: 5 Gaps: 6

US-10-757-093-4 (1-634) x BX363460 (1-1051)

QY 236 GlyLeuIleAsnTyrGluValGluValAlaAsnGlnThrThrglyGlnIleGlnIleSer 255  
 Db 1 GGGCTGGTGAATTCACATCTCTCTCAAGGAGTAACCTGTTCAAGTTGAGGTGGT 60  
 QY 256 ValIleAspGluAspGlyAlaIleValAlaLysAlaSerGlyAlaGlnGlyThrValThr 275  
 Db 61 CTTTGTGATGCAGAAACAAAGTCTGTGGCGAATGGAGCTGGAGCCAGGCCCAACTTAAG 120

QY 276 IleProSerValLysLeuTrpGlnPro-----GlyAlaAlaTyrLeuTyr 290  
 Db 121 GGGCCAGGTGTACAGCCTTGTGGCCGTACTGATGCACGAAGCCCTGCTTATCTGAT 180  
 QY 291 GlnLeuGlnValAsnIleValGly-----SerGlyAspValAlaAspThrTyrAsn 308  
 Db 181 TCATTGAGAGTGCAGCTACGACAGACAGACGTCACTGGGGCTGTGTGACTTCACTACACA 240  
 QY 309 LeuAlaThrGlyValArgThrValLysValAlaGlySerGlnPheLeuIleAsnGlyLys 328  
 Db 241 CTCCTGTGGGGATCCGACCTGTGGCTGTCCACAGACCCAGTCTCATATAGGGAAA 300  
 QY 329 ProPheTyrPheThrGlyPheGlyLysHisGlyLeuAspThrAlaValArgGlyLysGlyHis 348  
 Db 301 CTTTCTCTATTTCCACGGGTGTCAACAAGATAGATGGCAATCCGAGGAAAGGCTTC 360  
 QY 349 AspProAlaTyrMetValHisAspPheGlnLeuMetLysTrpIleGlyValAsnSerPhe 368  
 Db 361 GACTGGCCGCTGTGTGAAGACTTCAACTGCTTCGCTGGCTGTGGCCACAGCTTTC 420  
 QY 369 ArgThrSerHisTyrProTyrAlaGlnGluValMetAspPheAlaAspArgAsnGlyIle 388  
 Db 421 CGTACAGGCACTACCTCCATGACAGAGATGACAGATGTGACCGCTATGGGAT 480  
 QY 389 ValValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeuMetGlyValSerGlu 408  
 Db 481 GTGTCTATGATGATGATGTCGCCGCTGGCCCTGGCGCTG----- 519  
 QY 409 SerGlyAlaProGlnThrPheThrProAspAlaIleAsnAspLysThrGlnGluAlaHis 428  
 Db 520 -----CCGACAGTTCTTC-----AACACAGTTCTCTGCATCACAC 555  
 QY 429 LysGlnAlaIleArgGlnLeuIleAlaArgAspLysAsnHisAlaSerValValMetTrp 448  
 Db 556 ATGCAGGTGATGAAAGATGGTGGTGAAGACAAAGAACACCCCGGCTCGTATGGTGG 615  
 QY 449 SerIleAlaAsnGluProAlaSerHisGlyAspGlyAlaArgGlyLysPheGlnProLeu 468  
 Db 616 TCTGTGGCCACGAGCGTGTGCCACCTAGAAATGCTGCTACTTGAAGATGGTG 675  
 QY 469 ThrAsnLeuThrArgGlnLeuAspProThrArgProIleThrPheAlaAsnValGlyThr 488  
 Db 676 ATGCGTCAACACCAAAATCCTTGAGCCCTCCGCGCTGTGACCTTT-----GTGAGCAAC 729  
 QY 489 AlaThrTyrGlnLeuAspArgLysSerAspLeuPheAspValSerCysIleAsnArgTyr 508  
 Db 730 TCTAATCTATGCAGACAGACAGAGGGGCTCCGTATGTGATGTGATCTGTTGAACAGCTAC 789  
 QY 509 PheGlyTyrLysSerGlnThrGlyAspLeuGlnGluAlaGluAlaLeuGlnLysGlu 528  
 Db 790 TACTCTGTGATACAGACTACGAGGACCTGAGGTGATTCACTGCAGCCTGCCACCCG 849  
 QY 529 LeuHisGlyTyrGlnGlnLysPheHisArgProIleValMetThrGlnLysGlyValAsp 548  
 Db 850 TTGAGAACTGTGTATGAAGATATCAGAGCCCATTTATTCAGAGGAGATATGAGACAGA 909  
 QY 549 ThrLeu-AlaGlyLeuHisSerIleLeuGlyLeuProTyrSerGlnGluPheGlnValGlu 568  
 Db 910 ACGATTGGACGGTTTACACAGATCCACTGTGATCTGATCTGACAGAAAGATACAGAAAR 969  
 QY 568 MetLeuAspMetTyrHisArgValPheAspArgIleGlnSerMetAlaGlyGluHis-- 587  
 Db 970 TTYTCTAGACAGTACATCTGGGTGTGATTAATAAAMGMAAAATATGTGTGRRRCT 1029  
 QY 588 -ValTTP 589  
 Db 1030 CATTGG 1036

RESULT 9  
 CL486845/c 921 bp DNA linear GSS 01-APR-2004  
 LOCUS SAIL\_443\_B06.v2 SAIL Collection Arabidopsis thaliana genomic clone  
 DEFINITION



Db		493	TGTAATGTCTGGACAGCSTCAACCCGAATTACCATCAGCGAATCTTTGATGTCGTGCCT	434
OY		505	EAAATAAGTgThrPhgGIyTrpTySerGIInTHGLyASPLeuGlutJuaIagIuaIaIale	525
Db		433	GAAACCGTTATTAATTCGATGTATGTCCAAGCGCGCATTTGGAAAACGGCAGAAGACT	374
OY		525	UGLUyLysGIcUenHIGLyTrpGlnGluLyPheHISArgProIleValMetThrGluTy	545
Db		373	GGAAAAAGAACTTCTGGCCCTGGCAGAGAAACTGCATCAGCCGATTTATCATCACCGATA	314
OY		545	rgIyAlaAspThrLeuAlaGlyLeuHISerIleLeuGlyLeuDrOTPSerGIuGluPh	565
Db		313	CGGGGTGATACGTTAGCTACCGGGCTGCACATCATGTACACCGACATGTGAGTAGAAGTA	254
OY		565	eGlnValGImeCluaspMetyTrHISAArgValPheAspArgIleGusTermcAlaGI	585
Db		253	TCAGTGTCATGCGTGTGATATGTATCAACCGGCTTTGATGTCGCTCGCCGCCGTCTGG	194
OY		585	yGluHISvalTrpAsnPhelaaAspPheGlnTHASnLeuGlyIleIleArgValaAspGI	605
Db		193	TGAACAGGTATGGAAATTCGCCGATTTGGCACTTCGCAGAGGCATATTCGCCGTTGGCGG	134
OY		605	YASnLYSLeGIVaLPheThrARgAspArgLYSProLYSaLaalAHISserLeuAr	625
Db		133	TAAACAAGAAAGGATCTTCACTCGCGAACCGAAACGAGTCGGCGGTTTTAATGTAAg	74
OY		625	gaIa	626
Db		73	AAGC	70
RESULT_10				
CDS03098				
LOCUS				
DEFINITION			1138 bp mRNA linear EST 12-JUN-2003	
ACCESSION			CDAB60-D07.xid-t SHGC-CDA Gasterosteus aculeatus cDNA clone	
VERSION			CDAB60-D07 5', mRNA sequence.	
KEYWORDS			CD503098	
SOURCE			CD503098.1 GI:31433163	
ORGANISM			EST.	
			Gasterosteus aculeatus (three spined stickleback)	
			Gasterosteus aculeatus	
			Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
			Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;	
			Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;	
			Gasterosteidae; Gasterosteus.	
			1 (bases 1 to 1138)	
			Kingsley,D.M., Pelchel,C., Balabhadra,S., Grimwood,J., Dickson,M.,	
			Schmutz,J. and Myers,R.M.	
			Expressed sequence tags from Gasterosteus aculeatus	
			Unpublished (2003)	
			Contact: Kingsley, DM	
			HMMI and Department of Developmental Biology	
			Stanford University School of Medicine	
			Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA	
			Tel.: 650 725 5954	
			Fax: 650 725 7739	
			Email: kingsley@cgm.stanford.edu	
			Plate: 60	
			High quality sequence stop: 782.	
FEATURES				
Source				
			Location/Qualifiers	
			1..1138	
			/organism="Gasterosteus aculeatus"	
			/mol_type="mRNA"	
			/strain="Salinas river, CA"	
			/db_xref="taxon:69293"	
			/clone="CDA60-D07"	
			/sex="mixed male and female"	
			/tissue_type="heads and internal organs combined"	
			/dev_stage="adult"	
			/clone_1Id="SHGC-CDA"	
			/note="Vector: lambda ZAP Express/pBK-CMV; Site 1: EcORI	
			(5' adaptor); Site_2: XhoI (3' linker primer); The mixed	
			organ cDNA library was generated using the ZAP-cDNA method	
			by Stratagene. First strand cDNA synthesis was primed with	

a 50 bp linker primer containing an oligo dt sequence preceded by a synthetic XhoI site. 5 prime adaptors were used containing an EcoRI cohesive end. The finished cDNAs were inserted in to the ZAP express vector unidirectionally in the sense orientation with respect to the lacZ promoter of pBK-CMV. An amplified library was prepared from approximately 3 million primary clones in the lambda ZAP Express vector. In vivo excision was then used to generate individual pBK-CMV phagemid clones for EST sequencing."

## ORIGIN

## Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	770.50	166	47	111	25	5
Percent Similarity:	61.56%					
Best Local Similarity:	47.98%					
Query Match:	22.97%					

US-10-757-093-4 (1-634) x CD503098 (1-1138)

```

QY 284 ProGlyAlaAlaTyrLeuTyrGlnLeuGlnValAsnIleValGlySerSerGlyAspVal 303
Db 8 CCGAACCCAGGTTACCTTACTCTTTGGAGTGTCTTACTGCTGCTGAGGATCTGCT 67
QY 304 Val--AspThrTyrAsnLeuAlaThrGlyValAlaThrValAlaGlySerGln 322
Db 68 TCACACACACGTCGTCACCTTACCGACGTCGACGCTGACGCTTACACACACCG 127
QY 323 PheLeuIleAsnGlyLysProPheTyrPheThrGlyPheGlyLysIleAsnIle 342
Db 128 TTCTCTCAACAAACAAAGCCCTTCTTACTTCCACGAGTAATAAACAAGACGCTCGAT 187
QY 343 ValArgGlyLysGlyHisAspProAlaTyrMetValHisAspPheGlnLeuMetLys 362
Db 188 ATTGAGGCAAGAGCTGTCGACGTCGCTCATGTCGTCGTCGTCGTCGTCGTCGTCG 247
QY 363 IleGlyAlaAsnSerPheArgThrSerHisTyrProTyrAlaGluGluValMetAsp 382
Db 248 TTGGGGGCGCAACTGTTCCGACACACGACATACCTTATCAGAGAGATCTCGCAGATG 307
QY 383 AlaAspArgAsnGlyIleValValIleAspGluThrProAlaValGlyLeu--AsnIle 401
Db 308 TGTGACCGCGCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 367
QY 402 AlaLeuMetGlyValSerGluSerGlyAlaProGlnThrPheThrProAspAlaIleAsn 421
Db 368 CGCAGTTTGGG----- 379
QY 422 AspLysThrGlnGlnAlaHisIleGlnAlaIleArgGluLeuIleAlaArgAspLysAsn 441
Db 380 AACGCGCTTAAACCATCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 439
QY 442 HisAlaSerValValMetTTPSerIleAlaAsnGluProAlaSerHisGluLysGlyAla 461
Db 440 CATTCCCTGTGTGTCATGTGTGTCAGTGGCCAAAGAACCCGCTGACAGAGATGCTCTGCT 499
QY 462 ArgGlyTyrPheGluProLeuThrAsnLeuThrArgGlnLeuAspProThrArgProIle 481
Db 500 GATTACTATTTCAAACCTTGATMAAACATACCAAGAAATTGGATCCAAACCCGCGCGTC 559
QY 482 ThrPheAlaAsnValGlyThrAlaThrTyrGlnLeuAspArgIleSerAspLeuPheAsp 501
Db 560 ACTTTT-----ATCACAGACAGTAACTATGCCAGGATTAAGGGGCTCCCTACGTCGAC 613
QY 502 ValSerCysIleAsnArgTyrPheGlyTyrPyrSerGlnThrGlyAspLeuGluAla 521
Db 614 GTCATCTGTGTAACGATTACTCTCTGTCATCAGTACGACCCCGACCCGAGGTCATC 673
QY 522 GluAlaAlaLeuGluLysGluLeuHisGlyTyrGlnGluLysPheHisArgProIleVal 541
Db 674 CCCATCGACGTCACACCTCAAGTTTGAAGTACGTACGGAAGAGTACCAAGAACCCATCATC 733

```

```

QY 542 MetThrGluTyrGlyAlaAspThrLeuAlaGlyLeuHisSerIleLeuGlyLeuProTyr 561
Db 734 CAGAGCGAATATACGACGAGGATGGCGCCGCTTCAACAGTATCCACCCGATGTTT 793
QY 562 SerGluGluPheGlnValGlnMetLeuAspMetTyrHisArgValPheAsp-----Arg 579
Db 794 ACTGAGAGATACCAAGATTAATGTCCTGACAGACTTACCAACAGTGTTCACCAAGAAAGA 853
QY 580 IleGluSerMetAlaGlyGluHisValTyrAsnPheAlaAspPheGlnThrAsnLeuGly 599
Db 854 AG-CAGTACGTCATCGCGCAACTCTGGAAGTTCAGACTTATGACACACACAGG 912
QY 600 IleIleArgValAspGlyAsnIleLysGlyValAlaPheThrArgAspArgLysProLysAla 619
Db 913 ATCATCGGTCGTGTGGGAAACAGAG-GGTGTCTTCAGCAGGAAAGGCGCAAGCA- 970
QY 620 AlaAlaHisSerLeuArg 625
Db 971 GCAGCATTCATCTTAAG 988

```

## RESULT 11

## BX745933/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## source

1..857  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="TGA066d19"  
/dev\_stage="gastrula (stages 10.5-12 mixed)"  
/lab\_host="Escherichia coli XL1-blue"  
/clone\_id="XGC-gastrula"  
/note="Vector: PCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into PCS107 with EcoRI at the 5' end and NotI at the 3' end."

## ORIGIN

## Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:
Score:	1.9e-70	857	144	44
Percent Similarity:	726.50			
Best Local Similarity:	63.95%			
	48.98%			

Query Match:	21.668	Indels:	21
DB:	5	Gaps:	5
US-10-757-093-4 (1-634) x BX745933 (1-857)			



	BMS57676	1055 bp	mRNA	linear	EST-20-FEB-2002
LOCUS					
DEFINITION	AGENCOURT_6568423 NIH_MGC_98 Homo sapiens CDNA clone IMAGE:5477657				
ACCESSION	5', mRNA sequence.				
VERSION	BMS57676				
KEYWORDS	BMS57676.1	GI:18799844			
SOURCE	EST.				
ORGANISM	Homo sapiens (human)				
	Homo sapiens				
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 1055).				
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strusberg, Ph.D. Email: <a href="mailto:cgarbs-remail.nih.gov">cgarbs-remail.nih.gov</a> Tissue Procurement: ATCC CDNA Library Preparation: Rubin Laboratory DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <a href="http://image.lnl.gov">http://image.lnl.gov</a> Plate: LNCMI996 row: a column: 18 High quality sequence stop: 668.				
FEATURES	Location/Qualifiers				
source	1..1055				

REFERENCE  
1 (bases 1 to 877)  
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
AUTHORS  
Yi Y., Desai, R., Olarte, M., Henthorn, P. and George A.L.  
TITLE  
Expressed sequence tags from Canine heart  
JOURNAL  
Unpublished (2003)  
COMMENT  
Other ESTs: CH3#042\_G03T3  
Contact: George AL  
Division of Genetic Medicine  
Vanderbilt University  
529 Light Hall, 2215 Garland Avenue, Nashville, TN 37232-0275, USA  
Tel: 615 936 2660  
Fax: 615 936 2661  
Email: al.george@vanderbilt.edu  
Insert Length: 2032 Std Error: 0.00  
Seq primer: T7: TAATACGACTCACTATAGGCG  
High quality sequence start: 39  
High quality sequence stop: 859.  
Location/Qualifiers

FEATURES  
source  
1..877  
/organism="Canis familiaris"  
/mol\_type="mRNA"  
/db\_xref="taxon:9615"  
/clone="CH3#042\_G03"  
/tissue\_type="heart"  
/cell\_type="heart"  
/dev\_stage="mixed developmental stages (adult, 30 day - 40 day fetal)"  
/clone\_lib="Canine heart normalized cDNA library in pBluescript"  
/note="Organ: heart; Vector: pBluescript; Site 1: 5' of vector NOTI; Site 2: 3' of vector EORI; Tissue source: dog heart (adult, 30 day - 40 day fetal), right and left atria and ventricle. Dog breed - mixed (beagle, German shepherd, pointer, Irish setter). Library construction: oligo-dT primed"

## ORIGIN

Alignment Scores:  
Pred. No.: 1.54e-65 Length: 877  
Score: 683.00 Matches: 135  
Percent Similarity: 62.54% Conservative: 47  
Best local Similarity: 46.39% Mismatches: 87  
Query Match: 20.36% Indels: 22  
DB: Gaps: 3

US-10-757-093-4 (1-634) x CF406520 (1-877)

QY 192 G|YLSARGLIeGINThrYrGLnHIsAspPhETyZAsnTYrAlaG|YLeuAlaArgSer 211  
|||  
|||  
Db 46 GGTACTTCCTCCGACACACTACTTTCCTTCAACTACGCGGCGTCGATCGCCCT 105  
QY 212 ILeTPrLeuYrSerValProGlnGlnHIsIleGlnAspIleThrValValThraSpVal 231  
::: |||  
::: |||  
Db 106 GTGCTCCCTACACACACTACTTACCTACATCGACGACGACCGTCCACCGCGCTG 165  
QY 232 AspGlyAspArgnGlyLeuIleAsnTYrGluValAlaIleAsnGlnThrGlyGln 251  
|||  
|||  
Db 166 GACCAAGCACTGGGCTGTGATTCACGATTTTGTCCAGGCGCAGTGAACACTTCCAG 225  
QY 252 ILeGlnIleSerValIleAspGluAspGlyValAlaIleValAlaIleValAlaGln 271  
::: |||  
::: |||  
Db 226 CTGGAAGTATCTCTCTGATGAGGAGCAAGTCTGCGCCACGAGGAGCAGGAGCCAG 285  
QY 272 G|YThrValThrIleProSerValYsLeuThrGlnPro-----G|YAla 286  
|||  
|||  
Db 286 GGCCCGCTGCGAGTGGCCCAATGTCCACTGTGGTGGCGTACTGATGATGACGACCC 345  
QY 287 AlaTYrLeuTYrGlnLeuGlnValAsnIleValGly-----SerSerGlyAspValVal 304  
|||  
|||  
Db 346 GCTTACTCTGTACTCTCTGAGGTGAGGCTGACTGCGCAGATGCGCGCTGCTGTGCA 405  
QY 305 AspThrTYrAsnLeuAlaThrGlyValArgThrValYsValAlaGlySerGlnPheLeu 324  
|||  
|||

Db 406 GACTTATAACTCTCCCGTGGGATTCCGACCGTGGCCGCTCACAGAGCCCGACTTCTC 465  
QY 325 ILeAsnGlyYsProPheTYrPheThrGlyPheGlyYsHISGluAspThrAlaValArg 344  
|||  
|||  
Db 466 ATCAACGGGAAACCTTCTATTTCCATGGGCTCAACAACATGAGATGGCCGATATCCGA 525  
QY 345 G|YLYSG|YHIsAspProAlaTYrMetValHIsAspPheGlnLeuMetYsTrpIleGly 364  
|||  
|||  
Db 526 GGGAGAGGCTTTGACTGGCCGCTGCTGTGAAGACTTCAACCTGTGGCTGGCTGGGC 585  
QY 365 AlaAsnSerPheArgThrSerHISTYrProTYrAlaGluGluValIleMetAspPheAlaAsp 384  
|||  
|||  
Db 586 GCAATGCTTCCGACACGACCTACCCCTACCGGAGGAGTGAAGACACTCTGCGAC 645  
QY 385 ArgAsnGlyIleValValIleAspGluThrProAlaValGlyLeuAsnIleAlaLeuMet 404  
|||  
|||  
Db 646 CGCTATGGATCGTGTATGACGAGAGCCCTGTGTGGCATCATGCTGTC----- 699  
QY 405 G|YValSerGluSerGlyAlaProGlnThrPheThrProAspAlaIleAsnAspYsThr 424  
|||  
|||  
Db 700 -----CAGACTACAGCAATGTGTCC 720  
QY 425 GlnGluAlaHISGlyGlnAlaIleArgGluLeuIleAlaArgAspYsAsnHISAlaSer 444  
|||  
|||  
Db 721 CTGCAACACCACTGAGGTGATGGGAGCTGTGCTGGGATGAAGATCACCATCT 780  
QY 445 ValValMetTrpSerIleAlaAsnGluProAlaSerHISGluAspGlyAlaArgGluTYr 464  
|||  
|||  
Db 781 GTAGTCAATGTCTGTGAGCCCAATGAGCCCACTTCTTCTGAAAGCCCTGCTTACTAC 840  
QY 465 PheGluProLeuThrAsnLeuThrArgGlnLeu 475  
|||  
|||  
Db 841 TTCAAGAGCTGATTTGCTACACCAAGGCCCTT 873

Search completed: March 22, 2005, 21:57:16  
Job time : 4490 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: March 18, 2005, 23:42:44 ; Search time 43 Seconds

(without alignments)  
1418.638 Million cell updates/sec

Title: US-10-757-093-4

Perfect score: 3354

Sequence: 1 MKFLGLSLSLAAPSIGTP.....RKPKAAASHLRARMTSIDKN 634

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79: \*  
1: p1r1: \*  
2: p1r2: \*  
3: p1r3: \*  
4: p1r4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1711	51.0	603	1	GBEGCC	beta-glucuronidase
2	1338	39.9	651	2	A26581	beta-glucuronidase
3	1335	39.8	648	2	A25047	beta-glucuronidase
4	1323	39.4	648	2	A32576	beta-glucuronidase
5	1186	35.4	368	2	A85768	partial beta-D-glu
6	1186	35.4	370	2	D90919	beta-D-glucuronida
7	882	26.3	563	2	A72300	beta-glucuronidase
8	779	23.2	570	2	C90485	beta-glucuronidase
9	524	15.6	237	2	B85768	interrupted beta-D
10	524	15.6	237	2	B85768	partial beta-D-glu
11	426.5	12.7	1087	2	F72283	beta-galactosidase
12	408.5	12.2	1307	2	T35944	probable beta-gala
13	390	11.6	996	2	D86872	beta-galactosidase
14	389.5	11.6	1034	2	T30574	beta-galactosidase
15	382.5	11.4	1034	2	C83990	beta-galactosidase
16	376.5	11.2	1034	2	T30551	beta-galactosidase
17	372.5	11.1	1024	1	GBEC	beta-galactosidase
18	367	10.9	1025	1	JC1266	beta-galactosidase
19	359.5	10.7	755	2	D96842	probable beta-gala
20	355.5	10.6	1024	2	E90678	beta-D-galactosida
21	355.5	10.6	1024	2	A85529	beta-D-galactosida
22	353	10.5	1015	2	I39697	beta-galactosidase
23	342	10.2	1026	2	A49750	beta-galactosidase
24	332	9.9	1042	2	E85968	evolved beta-D-gal
25	332	9.9	1042	2	F91123	evolved beta-D-gal
26	331	9.9	1060	2	A10201	beta-galactosidase
27	329.5	9.8	897	2	A39405	beta-galactosidase
28	329	9.8	1042	1	GBEC	beta-galactosidase
29	325	9.7	1005	2	T31333	beta-galactosidase

30	323.5	9.6	626	2	A42891	beta-galactosidase
31	311.5	9.3	716	2	JU0275	beta-galactosidase
32	308	9.2	1075	2	T47603	beta-galactosidase
33	287	8.6	1007	2	A30093	beta-galactosidase
34	283.5	8.5	1034	2	A24925	beta-galactosidase
35	261.5	7.8	2228	2	E97942	beta-galactosidase
36	260.5	7.8	2233	2	B95075	beta-galactosidase
37	169	5.0	900	2	T19689	hypothetical prote
38	160.5	4.8	620	2	T37230	probable beta-mam
39	139.5	4.2	448	2	JM0038	beta-glucosidase (
40	139.5	4.2	785	2	H72228	hypothetical prote
41	126.5	3.8	448	2	T48214	endo-1,4-beta-mam
42	126.5	3.8	499	2	A27198	cellulase (EC 3.2
43	124.5	3.7	499	2	JN0111	cellulase (EC 3.2
44	124	3.7	891	2	AB0255	beta-mannosidase p
45	123.5	3.7	403	2	A84592	(1-4)-beta-mannan

#### ALIGNMENTS

RESULT 1  
GBEGCC  
beta-glucuronidase (EC 3.2.1.31) uidA - Escherichia coli (strain K-12)  
N/Alternate names: beta-D-glucuronoside glucuronohydrolase; gusa protein  
C/Species: Escherichia coli  
C/Date: 30-Jun-1988 #sequence revision 05-Dec-1997 #text\_change 09-Jul-2004  
C/Accession: C64918; 153717; A26487; S4355  
R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Coi  
A.; Rose, D.J.; Mau, B.; Shaoh, Y.  
A/Title: The complete genome sequence of Escherichia coli K-12.  
Science 277, 1453-1462, 1997  
A/Reference number: A64720; MUID:97426617; PMID:9278503  
A/Accession: C64918  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-603 <RES>  
A/Cross-references: UNIPROT: P05804; GB: A8000257; GB: U00096; NID: g178798; PIDN: AAC74689.J  
A/Experimental source: strain K-12, substrain MG1655  
R/Schlaman, H.R.; Risseuw, E.; Franke-van Dijk, M.E.; Hooykaas, P.J.  
Gene 138, 259-260, 1994  
A/Title: Nucleotide sequence corrections of the uidA open reading frame encoding beta-gl  
A/Reference number: 153717; MUID:94171050; PMID:8125312  
A/Accession: 153717  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-603 <RES>  
A/Cross-references: GB: S69414; NID: G545893; PIDN: AAB30197.1; PID: G545894  
R/Jefferson, R.A.; Burgess, S.M.; Hirsch, D.  
Proc. Natl. Acad. Sci. U.S.A. 83, 8447-8451, 1986  
A/Title: Beta-glucuronidase from Escherichia coli as a gene-fusion marker.  
A/Reference number: A26487; MUID: 87041472; PMID: 3534890  
A/Accession: A26487  
A/Molecule type: DNA  
A/Residues: 1-419; 'VHGNIS', 427-603 <JEF>  
A/Cross-references: GB: M14641; NID: g868017; PIDN: AAA68923.1; PID: g868020  
R/Punt, P.  
submitted to the EMBL Data Library, April 1994  
A/Reference number: S43555  
A/Accession: S43555  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1, 'V', 3-603 <PUN>  
A/Cross-references: EMBL: Z32701; NID: g475168; PID: g475169  
C/Comment: This acid hydrolase catalyzes the cleavage of a wide variety of beta-glucuron  
C/Genes: uidA  
A/Map position: 36 min  
C/Function:  
A/Description: catalyzes hydrolysis of beta-D-glucuronoside to D-glucuronate and alcohol,  
A/Pathway: starch and glucose metabolism  
C/Superfamily: beta-glucuronidase  
C/Keywords: glycosidase; hydrolase

Query Match 51.0%; Score 1711; DB 1; Length 603;  
 Best Local Similarity 55.0%; Pred. No. 6.1e-111;  
 Matches 330; Conservative 93; Mismatches 16; Indels 16; Gaps 9;

QY 40 VRPQTSRELVLNDGLMKFAL--ASGLNDTAPMTAPLPKGLCECPVASYNDIFISRE 96  
 DB 2 LRPEVETPREIKKLDGLMAFSLDRENGCIDQ--RWMSALQESALIAVPSGPNQFADAD 59  
 QY 97 IHDVGVWYVYOREVIVPKGSGQERLYVRAESATTHGRIYNNRNLVAHVGGYTPFEADV 156  
 DB 60 IRNAGVWVYOREVIVPKGMAQRIYLRFDVAVHYGVWVNNQDMHOGGYTPFEADV 119  
 QY 157 ELVAPGEKFLITGVNNELTHETIPPKITGTGNATGRIOTGYODFPVYAGLARSIMLYS 216  
 DB 120 PYVLAGSVRITVNNELMNGITIPPGVIT-DENGKKOSYDFPNYAGIHSVNLTY 178  
 QY 217 VPQOHIDITVYVY--DGDNGLIYEVAVANQTGQIQISVIDEGAIYAKASGAGTV 274  
 DB 179 TPNTWVDITVTVHADQCHNASVDQV--VAN--GDVVELRDADQGVATGQTSGL 234  
 QY 275 TIPSVKLMQGAALYLOLVNIVGSGDVDTVNLATGVTVKASQELINKPFTFG 334  
 DB 235 QVNVPHLMQREBYLYELCV--TAKSQTECDIYPLRVGISVAVKGQFLINHPFTG 292  
 QY 335 FGKEDTAVRGKGDPAVMYVDFOLMKIGANSFRTSHYPAEYVDFADNRGIVVDET 394  
 DB 293 FGRHEDADLRKGKGDVNLVMDHDLMDIGANSRISHYPAEEMDMADHEGIVVDET 352  
 QY 395 PAVGLNAL-MGVSESGAP-QTFPPDALINDKTOEAKQAIRELIAIDKNHASYVMMSIAN 452  
 DB 353 AAVGFNLISLIGFPAKNGKPELYSEEVNGETQCAHQAIKELIADKNHASYVMMSIAN 412  
 QY 453 EPASHEDGAREVEPELTNLTRQDPTRPITFANVTATYQDLRIISDFVSCINRFGWY 512  
 DB 413 EPDTRPGAREYRPLAEATRKLDPTRPITCVNMFCDANTDTISDFVLCNRYGMY 472  
 QY 513 SQTGDEEAPALKEKELHMQEKFHRPIVWTEYGADTLAGLSILGPMSEEFQVOMLDM 572  
 DB 473 VQSGDLETAKEVLEKELIAWQEKHQPIITTEYGVDTLAGLSMYDMSEEFQCAMLDM 532  
 QY 573 YHRVPRISMAGEHWNFADFQTNLGIIVNDKKGVPFRDRKPKYAAHSLPARWTSID 632  
 DB 533 YHRVPRISAVBQVNNFADFATISQILRVGNKGIIFTRDRKPKSAFLIQKRWTKMY 592

RESULT 2  
 A26581  
 beta-glucuronidase (EC 3.2.1.31) precursor - human  
 C/Species: Homo sapiens (man)  
 C/Date: 05-Oct-1988 #sequence, revision 05-Oct-1988 #text\_change 09-Jul-2004  
 C/Accession: A26581; A40337; A24983; A36538  
 R/Oshima, A.; Kyle, J.W.; Miller, R.D.; Hoffmann, J.W.; Powell, P.P.; Grubb, J.H.; Sly,  
 Proc. Natl. Acad. Sci. U.S.A. 84, 686-689, 1987  
 A/Title: Cloning, sequencing, and expression of cDNA for human beta-glucuronidase.  
 A/Reference number: A26581; MUID:87118233; PMID:3468507  
 A/Accession: A26581  
 A/Molecule type: mRNA  
 A/Residues: 1-651 <OSH>  
 A/Cross-references: UNIPROT:P08236; GB:M15182; NID:g183232; PIND:AAA52561.1; PID:g183233  
 A/Experimental source: placenta  
 R/Shikey, U.M.; Miller, R.D.; Wu, B.M.; Grubb, J.H.; Christensen, S.G.; Kyle, J.W.; Sly  
 Genomics 10, 1009-1018, 1991  
 A/Title: Analysis of the 5' flanking region of the human beta-glucuronidase gene.  
 A/Reference number: A40337; MUID:92009900; PMID:1916806  
 A/Accession: A40337  
 A/Molecule type: DNA  
 A/Residues: 1-70 <SHI>  
 A/Cross-references: GB:M65002; NID:g183706; PIND:AAA52622.1; PID:g183707  
 R/Guise, K.S.; Korneluk, R.G.; Waye, J.; Lamhonwah, A.M.; Quan, F.; Palmer, R.; Ganschow  
 Gene 34, 105-110, 1985  
 A/Reference number: A24983; MUID:85232043; PMID:3924735  
 A/Accession: A24983

A/Molecule type: mRNA  
 A/Residues: 520-585 <GUT>  
 A/Cross-references: GB:M10618; NID:g183704; PIND:AAA52621.1; PID:g183705  
 R/Tomatsu, S.; Fukuda, S.; Sukegawa, K.; Ikeda, Y.; Yamada, Y.; Sasaki, T.; (C  
 Am. J. Hum. Genet. 48, 89-96, 1991  
 A/Title: Mucopolysaccharidosis type VII: characterization of mutations and molecular hete  
 A/Reference number: A36538; MUID:91090114; PMID:1702266  
 A/Accession: A36538  
 A/Molecule type: mRNA  
 A/Residues: 378-385, 616-621, 643-651 <TOM>  
 A/Genetics:  
 A/Gene: GDB:GUSB  
 A/Cross-references: GDB:120025; OMIM:253220  
 A/Map position: 7q22-7q22  
 C/Superfamily: beta-glucuronidase  
 C/Keywords: glycoprotein; glycosidase; homotetramer; hydrolase; lysosome  
 F/1-22/Domain: signal sequence #status predicted <Sig>  
 F/23-651/Product: beta-glucuronidase, placental #status predicted <Mat>

Query Match 39.9%; Score 1338; DB 2; Length 651;  
 Best Local Similarity 44.3%; Pred. No. 5.7e-85;  
 Matches 274; Conservative 98; Mismatches 199; Indels 48; Gaps 12;

QY 42 PORTSRELVLNDGLMKFALASGLNDT-----AQPMTAPLPK--LECPVASYNIF 92  
 DB 30 PQESPBECKELDGLMSF--RADPSDNRRGEEQYRPLMESGPTVDMPPSSFNDIS 87  
 QY 93 ISREIHDVGVYVYOREVIVPKGSGQ--RYVRAESATTHGRIYNNRNLVAHVGGYT 149  
 DB 88 QDMRLHFPVGVWVYVREVLVPERMTQDTRTVLIRIGASHAYIAVANGVDTLEHGYYL 147  
 QY 150 PFEADVTELVANG--EKRLITGVNNELTHETIPPKI-----TTGNATGRIOTYQHD 201  
 DB 148 PFEADISNLVQVGRPSRRLITLNTLPTLPBGTQYITDSKPKGYVQNTYFD 207  
 QY 202 PNYVAGLARSIMLYSVPOOHIDITVWVDVDGNGLIYEVAVANQTGQIQISVIDEG 261  
 DB 208 PENVAGLQSVLTYTPPTIYIDITVTSVQDSGLVNYQISVKSNNLFKLEVRLLDMEN 267  
 QY 262 ALVAKASGAGTVITPSYKLMQP-----GAALYLOLVNIVG--SSGDVDTVNLATGVR 314  
 DB 268 KVVANGTGTGGQKVPYGLWVPMYLMHRPAILVLSLEVQTLAQTSLGVSDFYTLPGVIR 327  
 QY 315 TVKVAGSOFLLNGKPFYTGFGKEDTAVRGKGDPAVMYVDFOLMKIGANSFRTSHY 374  
 DB 328 TVAVTKSQFLNGKPFYTHGVNKKEDADIRKGFDPMLVVDQFNILRWLGNAPFTSHY 387  
 QY 375 YAEVYDFADNRGIVVVIDETPAVGLNALMGVSESGAPQTFPPDALINDKTOEAKQAI 434  
 DB 388 YAEVYQMCNDRGIVVIDECFQVGLAL-----PQF-----NNVSLHHMQVWEE 432  
 QY 435 LIARKNHASYVMMSIANEPASHEDGAREVEPELTNLTRQDPTRPITFANVTATYQD 494  
 DB 433 VVRKRNHAPVAVMSVNAEPASHESAGYILKVIYAHKSLDPRPVTF--VSSNSVAAD 490  
 QY 495 RISLDFVSCINRFGWVSQTGDEEAPALKEKELHMQEKFHRPIVWTEYGADTLAGLI 554  
 DB 491 KGAPVVDVICANSYSYSWHDVGHLELLOLATQTPENWKKYQKPIIOSEVGAETLIGFH 550  
 QY 555 SILGPMSEEFQVOMLDMYHVRFD--RIESNAGEHWNFADFQTNLGIIVNDKKGKVPFT 612  
 DB 551 QDPRLMFEEYQKSLLEQYHGLDQKRRKYVAGELINWVADFMTEQSPTRVLGNKKGIFT 610  
 QY 613 RDRKPKAAHSLPARWTSI 631  
 DB 611 KQRPKSAFLIRERYWKI 629

RESULT 3  
 A25047  
 beta-glucuronidase (EC 3.2.1.31) precursor - rat  
 C/Species: Rattus norvegicus (Norway rat)  
 C/Date: 30-Jun-1988 #sequence, revision 30-Jun-1988 #text\_change 09-Jul-2004



C/Accession: A25047; S00345  
 R.Nishimura, Y.; Rosenfeld, M.G.; Kreibich, G.; Gubler, U.; Sabatini, D.D.; Adenik, M.;  
 Proc. Natl. Acad. Sci. U.S.A. 83, 7292-7296, 1986  
 A/Title: Nucleotide sequence of rat preputial gland beta-glucuronidase cDNA and in vitro  
 A/Reference number: A25047; PMID:87016933; PMID:3463967  
 A/Accession: A25047  
 A/Molecule type: mRNA  
 A/Residues: 1-648 <RNA>  
 A/Cross-references: UNIPROT:P06760; GB:M13962; NID:g204329; PIDN:AAA41228.1; PID:g204330  
 A/Experimental source: Female preputial gland  
 R/Powell, P.P.; Kyle, J.W.; Miller, R.D.; Pantano, J.; Grubb, J.H.; Sly, W.S.  
 Biochem. J. 250, 547-555, 1988  
 A/Title: Rat liver beta-glucuronidase: cDNA cloning, sequence comparisons and expression  
 A/Reference number: S00345; PMID:88183378; PMID:3355537  
 A/Accession: S00345  
 A/Molecule type: mRNA  
 A/Residues: 'E', 15-20, 'L', 22-486, 'L', 488-648 <POW>  
 A/Cross-references: EMBL:Y00717; NID:956270; PIDN:CAA68705.1; PID:g758260  
 C/Superfamily: beta-glucuronidase  
 C/Keywords: glycosidase; hydrolase  
 F:1-22/Domain: signal sequence #status predicted <SIG>  
 F:23-648/Product: beta-glucuronidase #status predicted <MAT>

Query Match 39.4%; Score 1335; DB 2; Length 648;  
 Best Local Similarity 44.3%; Pred. No. 9.2e-85;  
 Matches 275; Conservative 96; Mismatches 200; Indels 44; Gaps 12;

QY 42 PQTSSRELVLNGLMKFALASGLNDTAP-----WTAPPKG---LECVPSYNDIFI 93  
 DB 30 PKETPRLRLVLDGLMFSR-ADYSNNRLQGEKQWQRLREGPPTLDMVPSFNDITQ 88  
 QY 94 SREIHVGVVYQREIVPQKWSOE---RYLVRAESATHGRIYVNNRLVAHVGVGYP 150  
 DB 89 EAEIRNFIVGWYEREVLQKRTQDDRVVLEINSAHYAYVAVVGVGVHVEGHLR 148  
 QY 151 FEADVTELVAPG--EKERLTGVNNELTHEITPGKIT-----TGNATGRIQTYQDFY 203  
 DB 149 FEADIRKLVGSGPLTFRTVIAINNTLTPTLPRTGTVKMTDSMKYGFVQDISDF 208  
 QY 204 NVAGLARSLVYVPOOHIDITVVDGNGLIINTEVEVANNQTOGQIDISVIDEGAI 263  
 DB 209 NVAGLRSLVYVLTPTTYIDITVTDVDRDVGIVNWMISVQSGDHFQLEVRLLDEGKI 268  
 QY 264 VAASAGQVTITPSVTLMP-----GAATLQGVNIVSSSGDVDTVTLANGVTVKY 318  
 DB 269 VAGTGNESGLKVPRAHLMPTLMEHPALYLSLEVMT--TPSSVSDFYTLVPGIRVAV 327  
 QY 319 AGSQPLINGKPFYFTGFKHEDTAVRGKHDPAMVMDPOLMKVIGANSFRTSHYPAEE 378  
 DB 328 TKSFLINGKPFYFGVGNKHEDSIRGRFDWPLIIDFPLMLKLGANSFRTSHYPAEE 367  
 QY 379 VMDPADNGIVIDEFPAVGLNALMGVSESGAPQFTTPAINDKTOEAKHQAIRELIAR 438  
 DB 388 VLQLCRYGIVVIDECGVGIVL-----PQSF-----GNVSLRHLLEVMDLVRR 432  
 QY 439 DKHNASVYVMSINERPSHEDGAREVEEPLTNLTRODPTPTFANVATATQOLDISD 498  
 DB 433 DKHNASVYVMSINERPSHEDGAREVEEPLTNLTRODPTPTFANVATATQOLDISD 490  
 QY 499 LPDVSCINRYFGVYSGTGLEBAALAEKELHGOEKFHRRPIVWTEGADTLALHSLILG 558  
 DB 491 YVDVICNSLWYHDGHEVLQQLTSQFEMMYKMYKQKPIIOSEVGAALVSGLHDDP 550  
 QY 559 LPMSEFQVOMDMYHVRVD--RISNAGHVNWVADPQNLGIIRVGNKKGVFTDRK 616  
 DB 551 RMSEFEQVQTLNRYHLIDKEKREYVIGELINWVADPQNLGIIRVGNKKGVFTDRK 610  
 QY 617 PKAASLRARMTSI 631  
 DB 611 PKAASLRARMTSI 625

RESULT 4

A32576  
 beta-glucuronidase (EC 3.2.1.31) allele B precursor - mouse  
 N/Alternate names: beta-D-glucuronoside glucuronohydrolase  
 C/Species: Mus musculus (house mouse)  
 C/Date: 12-Oct-1989 #sequence, revision 12-Oct-1989 #ext.change 09-Jul-2004  
 C/Accession: A32576; B32576; I49692; A28954; A29977; A35798  
 R/Mawzyrak, P.M.; Gallagher, P.M.; D'Amore, M.A.; Carter, J.E.; Lund, S.D.; Rinchik, E.  
 Mol. Cell. Biol. 9, 4074-4078, 1989  
 A/Title: DNA determinants of structural and regulatory variation within the murine beta-g  
 A/Reference number: A32576; PMID:89384641; PMID:2779578  
 A/Accession: A32576  
 A/Molecule type: mRNA  
 A/Residues: 1-648 <MA>  
 A/Cross-references: UNIPROT:P12265; GB:M28540; GB:M27816; NID:g193718; PIDN:AAA63307.1; I  
 A/Experimental source: allele B  
 A/Accession: B32576  
 A/Molecule type: mRNA  
 A/Residues: 1-86, 'I', 88-648 <MA>  
 A/Cross-references: GB:M28541; NID:g193720; PIDN:AAA63308.1; PID:g193721; GB:M27816  
 A/Experimental source: allele H  
 R/Punkstein, B.; Leary, S.L.; Stein, J.C.; Catterall, J.F.  
 Mol. Cell. Biol. 8, 1160-1168, 1988  
 A/Title: Genomic organization and sequence of the Gus-s-a allele of the murine beta-gluc  
 A/Reference number: I49692; PMID:88216590; PMID:2835664  
 A/Accession: I49692  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-232, 'T', 234-264, 'D', 266-319, 'V', 321-427, 'K', 429-615, 'L', 617-648 <RS>  
 A/Cross-references: GB:M19279; NID:g193524; PIDN:AAA31697.1; PID:g309257  
 A/Experimental source: allele A  
 R/D'Amore, M.A.; Gallagher, P.M.; Korfhagen, T.R.; Ganschow, R.E.  
 Biochemistry 27, 7131-7140, 1988  
 A/Title: Complete sequence and organization of the murine beta-glucuronidase gene.  
 A/Reference number: A28954; PMID:89062453; PMID:3196706  
 A/Accession: A28954  
 A/Molecule type: DNA  
 A/Residues: 1-264, 'D', 266-319, 'V', 321-648 <DM>  
 A/Cross-references: GB:J02836; GB:J03035; GB:M20204; NID:g193716; PIDN:AAA98623.1; PID:g  
 R/Gallagher, P.M.; D'Amore, M.A.; Lund, S.D.; Ganschow, R.E.  
 Genomics 2, 215-219, 1988  
 A/Title: The complete nucleotide sequence of murine beta-glucuronidase mRNA and its deduc  
 A/Reference number: A29977; PMID:88284700; PMID:3397060  
 A/Accession: A29977  
 A/Molecule type: mRNA  
 A/Residues: 1-264, 'D', 266-319, 'V', 321-648 <GAL>  
 A/Cross-references: GB:J03047; NID:g193522; PIDN:AAA37696.1; PID:g309256  
 R/Li, H.; Takeuchi, K.H.; Manly, K.; Chapman, V.; Swank, R.T.  
 J. Biol. Chem. 265, 14732-14735, 1990  
 A/Title: The propenide of beta-glucuronidase. Further evidence of its involvement in cor  
 the serpin superfamily.  
 A/Reference number: A35798; PMID:90368633; PMID:2394651  
 A/Accession: A35798  
 A/Status: not compared with conceptual translation  
 A/Molecule type: mRNA  
 A/Residues: 593-648 <LIR>  
 A/Note: the location of the propenide cleavage site was not demonstrated directly but r  
 C/Comment: In some tissues, a portion of this enzyme is retained in the endoplasmic retic  
 syn.

C/Genetics:  
 A/Map position: 5  
 A/Intons: 70/3; 132/3; 193/2; 241/1; 303/3; 351/3; 411/2; 460/2; 488/3; 547/3; 593/1  
 C/Superfamily: beta-glucuronidase  
 C/Keywords: glycosidase; hydrolase; lysosome  
 F:1-22/Domain: signal sequence #status predicted <SIG>  
 F:23-648/Product: beta-glucuronidase, ER-retained form #status predicted <ERMT>  
 F:23-633/Product: beta-glucuronidase, lysosomal (default) form #status predicted <LMA>  
 F:634-648/Domain: carboxyl-terminal propenide #status predicted <CPRO>

Query Match 39.4%; Score 1323; DB 2; Length 648;  
 Best Local Similarity 44.3%; Pred. No. 6.3e-84;  
 Matches 272; Conservative 99; Mismatches 201; Indels 42; Gaps 12;

QY	42	PORTSRELVNLDGIMKFL---ASGLNDTQOPW-TAPLPG---LECPVASTNDIFLS	94
Db	30	EKESRSRLKALDGLMHFRADLSNNRLOGFQEDQYROPRLRSQFVLDPSPFSFIDQOE	89
QY	95	REIHDHVMYVYOREVIVPKMSOE---RYIVRASATHGRIYVNNRLVAEHVGYRPF	151
Db	90	AALRDFIGMVMYEREAALPRMTQDTRVVLIRINSAYVAVVWNGIHVEHGGHLPF	149
QY	152	EADYTELVAFG--EKFRLTIGVNNELTHETIPRGL-----TTGNATKRIQTYQHDPYN	204
Db	150	EADISKLVOQSPFLTTCRIITAIINMTLPHLTPPGTIVYKTSDTSWPKCYFQDTSFDFEN	209
QY	205	YAGIARSIWLSVSPQOHIQDITVTVTDVGDGNGLIINEVEVANQTTGOISVIDEDGAIV	264
Db	210	YAGIHRSVVLVLTPTTYITDITVITINVEODIGLVITWISVQSEHFOLEFVLDBEGKAV	269
QY	265	AKASGAGCTVTIPSVYKMQP-----GAAYLYOLOVNIAGSSGDVVDYTNALTVRITYKVA	319
Db	270	AHGNGOGOLOVPSANLWMPYLMHEHPAYMSLEVKVT--TSESVTDYVTLPIGIRTAVVT	328
QY	320	GSOFINKRPYFPGFGHEDTAVRGKGDHDAVWHDQMLKMGANSFRSHPYAEV	379
Db	329	KSKFLINKRPYFQGVNGBHSDIRKGGEFMDPLVVKPNLIRMGANSFRSHPYAEV	388
QY	380	MDPADRNGIVIDETPAVGNIALMGVSESGAPOTFETDAINDKTOEAKOARELIARD	439
Db	389	LQLCDRGIVVIDECPGVIVL-----PQSGNESLRH-----HLEVMELVARD	433
QY	440	KNHASVYVWMSIANEPASHEDGARREYEPRLNLTQOLDPTREITRANGTATYQOLDRIIDL	499
Db	434	KNHAVVYVWMSVANSPPSSALKPAAVYFKTLIHTHALDLTREVT--VSNACYDADLCLAPY	491
QY	500	FDVSCINRPFQWYSQOTGDLFEABALEKEKELHGMQOKFHRPIVMEYEGADTLAGHSITGL	559
Db	492	VDVVCANSYFBNYDYGHLEVIQOLNSQFENWYKTHOKPILIOSEYIGADAI PGHEDPPR	551
QY	560	PWSEEFQVOMLDMYHRVFD--RIESMAGEHWNEFADPQNLGIRVGNKKGVFTTRDKP	617
Db	552	MFSEEYQCAVLENHSHVLDQSRKEVYVVGELIWNFADFNQNSPLRVLGNKKGIFTRQOP	611
QY	618	KAAHSLPARWTST 631	
Db	612	KTSAPILIRERYWRI 625	

```

RESULT 5
A:58768
partial beta-D-glucuronidase [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: A85768
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.T.; Davis, N.W.; Lim, A.; Dimalante, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85460; MUID:21074935; PMID:11206531
A:Accession: A85768
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-368 <STO>
A:Cross-references: UNIPROT:O8K671; GB:AE005174; NID:912515602; PIDN:AMG56605.1; GSPDB:G
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
;Gene: uidA_1

```

	Query Match	35.4%	Score 1186	DB 2	Length 368
	Best Local Similarity	63.3%	Pred. No. 8.6e-75		
	Matches	224	Conservative 41	Mismatches 85	Indels 4
					Gaps 3
QY	281	LMQGAAYLYLQVNIIVSSGSDVDYTNLALGVRYKVASGQFLNGRPFFTGCKED			340
		LMQGEGLVEHYLY--TAKSRTECIYPLRVGIRSAVKQGFLLNHRPFPTGGRGD			63

Qy	341	TAVRGKHDAVWVHPOQLMKITGANSFPRSHVPAVEEMDPADRGIVIVETPAVGN	400
Qy	:		:
Db	64	ADJGKGFVDVILVHDLALMDVIGANSYRSHVPAVEEMIDMDEHGIVIVIDTAIVGN	123
Qy	401	IAL-MGVSESGAP-QFTTPDAINDKTOEAHKAQIRELIARDKXHASVVMMSIANEPASHE	458
Db	124	LSJGIGFEANCKPEKELVSEAVNGETQOAHLOAIKELIARDKXHPFVVMMSIANEDTPR	183
Qy	:		:
Qy	459	DGAREYEPFLTNLTROLDPRPITTFANVGATVQOLRISDLFPVGSCTNNYFGYSGTGL	518
Db	184	QGAREYAPFLAEKTRKLDPRPITTCVNMFCDAHTTIDSLDFVCLNNTYGYVOSGDL	243
Qy	519	EEAEPALEKELHMOEKFRPPIVMTVEYGADTLAHLISLGLPWSEEFQVOMLDMYHRVD	578
Db	244	ETAEKVLKELLMOEKLHPILITEYGVDTLAGLSMTDWMSESEYQCAWMLDMYHRVD	303
Qy	579	RISAMGEHVMFADFCOTNLGIRVDGNKKGVTPRRKRKAASHSRAPWTSID	632
Db	304	RVSANVGEQVNMADPFAITSQILRVGANKKGITTRRKRKSAAPFLLOKKNITGMN	357

```

RESULT 6
D90919
beta-D-glucuronidase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)
C|Species: Escherichia coli
C|Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #ext_change 09-Jul-2004
C|Accession: D90919
R|Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.G.,
Gasawara, N., Yasunaga, T., Kohara, S., Shiba, T., Hattori, M., Shinagawa, H.
DNA Res. 8, 11-22, 2001
A|Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome
A|Reference number: A98629, MUID:21156231, PMID:11258796
A|Accession: D90919
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-370 <NAV>
A|Cross-references: UNIPROT:Q8X671; GB:BA000007; PIND:BA83547.1; PID:g13361791; GSPDB:GM05952
A|Experimental source: strain O157:H7, substrain RIMD 0509952
C|Genetics:
C|Gene: EC82324

```

Query Match	35.4%	Score 1186	DB 2	Length 370
Best Local Similarity	63.3%	Pred. No. 8.7e-75		
Matches 224	Conservative 41	Mismatches 85	Indels 4	Gaps 3
QY	281	LMQGAAYLYOLQVNVIVGSSGDVDPDYNLATVRITVKAGSOFLLNGKPFYTGCKIED	340	
Db	8	LMQPGSGVLYELV--TAKSRTECDIYPLRVGIRSAVAVGEOFLNNKRFYTGGRHED	65	
QY	341	TAVRGKGHPAYVHDFOLMKWIIGANSFRTSHYPAAEVMDPADKNGIYVIDETAVGIN	400	
Db	66	ADRRGGPFNVLMVHDLMDWIGANSYRTSHYPAAEMLDMADRHGIVVIDETAAGFN	125	
QY	401	IAL-MKVSESGAP-OFTPDALINDKQEAHKQAIRLIRADGNHSAVVMWSIANEPASGE	458	
Db	126	LSIGIFEGANKPKELYSSEAVNGETQQAHLQAIKELLIRADKNHSSVVMWSIANEPDTRP	185	
QY	459	DGAREYFEBPLNLTROLDPREITTFANVTATYOLDRISLDFVASCINRPFQWYSQTDL	518	
Db	186	QGAREFYAPLAEBATRLDPTTRPTTCVNVWFCDAHDTDISLDFVCLNRYYGWYQSDLL	245	
QY	519	EEEBAALEKEHLHGWQKCFHRPIVMTEYGADTLAIGLSILGLPWSEEFQVQMLDMYHAFD	578	
Db	246	ETREKYLEKEELLAMOEKLOPIIITEYGVDTLAIGLSHMTWDMWSESEYCAMLDMYHAFD	305	
QY	579	RIRSMAGEYVWNPADFQTLGIIRVDGNKKGVFTDRPKAAASHLRARMTSID	632	
Db	306	RYSAAVGEQVWNPADPATISQGLIRVGNKKGIETTRDRPKSAALFLIQKRWGMAN	359	

RESULT 7  
A72300  
beta-glucuronidase - *Thermotoga maritima* (strain MSB8)







C:Genetics:  
A:Gene: bgam  
C:Superfamily: beta-galactosidase

Query Match	11.6%;	Score 389.5;	DB 2;	Length 1034;
Best Local Similarity	21.0%;	Pred. No. 9,4e-19;		
Matches 154;	Conservative 104;	Mismatches 261;	Indels 213;	Gaps 28;
QY	14	APSLGTFAARHPRNMTQHECPFLIKVRPQRTSRELNV-----LDGLMKFALAS	63	
DB	12	APANGTPEKNNNDPEITQLNRKSHALLMYQYVEELKNDKRSVYVQSGNSWYFHPAE	71	
QY	64	GLNDTAQPTAP-----LP-----KGLIECP-----	VPASY	88
DB	72	NADGRVKNFPAPFSEYKMDISVPSHMOLOGDYDQYTNVYVWVNELEPPAPYKY	131	
QY	89	NDIFISREIHDHWYVYQREYVVPKWSGERLYVAESATHHGRLYNNRLVAEHVGY	148	
DB	132	NPV-----GQYVFTTPKSEWMDQPYIFSGVESAFYWMGEFVGSSEDF	179	
QY	149	TPEADVTTELVAPEKFRLLTIGVNNELTHETIPPGKITGNATGKRIQYQ-----	199	
DB	180	TPAEFDITSLQGEN-----TIAV-----	EYRMSDASMLBD	212
QY	200	HDFYVYAGLARSIMLYSVPOOHIDITVTVDVG--DNGLINVEVANQTTGQIQ----	253	
DB	213	QDFWRMSGIFRDVYLYSTPQVHIYDFSVRSLLNNYEDGELVSADILNFEHDTDLTP	272	
QY	254	-ISVIDDGAIVAKAS-----GAQGTVT-----IPSVKLMGPAAYLYQLOVNI	VGSSG	301
DB	273	EVMILYDAAQOVLQAPLOTLSVSDQRTVSLRTHIKSPAKMSAESPLVLYLVLSTKMAAG	332	
QY	302	DVADVTYMLATGRTVTVKAGSQFLNGKPYFTGFGKHEDTAVRGK--GHDPAYVWHPOLM	360	
DB	333	SIITESCCKGFRFEIKNGIMTNGKRIYLRGVNRHEFDSYKGRAGITREDMIHILM	392	
QY	361	KWIGANSFRISHYAAEWDADFNRNGIVIDEPANGLNIALMGVSESGAPOTTPDAI	420	
DB	393	KOHININAVRTSHYNDVSWYELCNEYGLYVIDE-----NLETHG-----	TWT-----	435
QY	421	NDKQEAHKQAI-----RELARDKNHSAVVMMSIANEPASHEDGAREYF	465	
DB	436	--YIOGEQKAVPSKPEWKENVLDRCRSMYERKNNPSIIIMSGNESFGGEN-----	F	488
QY	466	EPLTNLRQLDPTPTTFPANGATYQLDRIISDLFDVSCINRYGWSQGDLEEAAL	525	
DB	489	QHWYTFEKKDSTRLVHEGI-----FHHRD-----YDASDIESTN--YVKPADVERVALNN	538	
QY	526	EKEIHGWEKHPRIVMTVEG--ADTLAHLISILGLPWSHEFQVOMLDMYHRVFRIES	582	
DB	539	PK-----KPYTLCEYSHAMNGSCGNLYK-----	YMWELFDQYPI	571
QY	583	MAGEHVNMFAD-----FQTNLGIIRVDGN--KKGVFTRD--RKPKAAS	623	
DB	572	LQGGIWMWMOQAQATAGDGTSYLAGDGFDPNDGNFCGNGLIPADGTASPKIAEVK	631	
QY	624	---LRARTSID	632	
DB	632	KCYCPVKMTAVD	643	

RESULT 15

C83990  
beta-galactosidase BH2723 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004  
C:Accession: C83990  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20512582; PMID:11058132  
A:Accession: C83990  
A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-1014 <STO>  
A:Cross-references: UNIPROT:Q9K9C6; GB:AP001516; GB:BA000004; NID:G10175192; PIDN:BA0644  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH2723  
C:Superfamily: beta-galactosidase

Query Match	11.4%;	Score 382.5;	DB 2;	Length 1014;
Best Local Similarity	21.9%;	Pred. No. 2.8e-18;		
Matches 144;	Conservative 97;	Mismatches 253;	Indels 163;	Gaps 25;
QY	52	NLDGLMKFALASGLNDTAQPTAP-----LPKGLIECPVASYNDIFISREIH-----DHV	101	
DB	47	SLNGHMKETHA-----INPWRPKREFYQGFDC--KCWDDILVPGHILOQGYGKPYV	97	
QY	102	GMVY-----YQREIVYKWSGERLYVRAASATHHGRLYNNR	139	
DB	98	NWYFWDGHHHLRPELPEDDNPVGSYKYPDIPNNNSNHELFISFGVETAFYVWLNGE	157	
QY	140	LVAEHVGCTPEPADVTTELVAPEKFRLLTIGVNNELTHETIPPGKITGNATGKRIQYQ	199	
DB	158	FVGSSEDSFTPAEFDITPYAVEG-----NKLCEVYQ-----RSTGSMLE--D	199	
QY	200	HDFYVYAGLARSIMLYSVPOOHIDITVTVDVG--DNGLINVEVANQTTGQIQ--S	255	
DB	200	QDFWRFSGIFRDVYLYTPIHIVYDMHVRADLDRSLQTLGILETLLEKRSQEKEMIVAE	259	
QY	256	VIDEDGAIYA-----KASAGQGTV--TIPSVKLMGPAAYLYQLOVNI	VGSSG	309
DB	260	LYDAAGAVAVATADMTNDQATVSNVSVPALMSAEDPYLYKLPLKLPDENGLIVEVVPQ	319	
QY	310	ATGVRTVTVKAGSQFLNGKPYFTGFGKHEDTAVRGKGHDPAYVWHPOLMKWIGANSFR	369	
DB	320	KIGFRFELVNNIMTLNGKRIYLFKGVNRHEFNRGRTGVVTKEDMLDIKTKKKNINAVR	379	
QY	370	TSHYPAEWDADFNRNGIVIDETPAVGLNIALMGVSESGA-----PQTFPPDAIN	421	
DB	380	TSHPNNSWYQLOCEYGLYVIDE-----NLETHGSMQKGLVSPSWNIPGNHL	429	
QY	422	DKQEAHKQAIELARDKNHSAVVMMSIANEPASHED--GAREYFEPLTNLRQLDPTR	479	
DB	430	EWEPYIMDRAV--SMEPRDNHPSIIMSGNESYAGEYILAVSRF-----KSVDSR	481	
QY	480	PITFANVGATYQLDRIISDLFDVSCINRYGWSQGDLEEAALAEKELHGW--QEKHR	538	
DB	482	LVHVEGVFHARA-----YDATSDMESRMVAKPKDIEDYLTNDPRK	521	
QY	539	PVMTVEY--GADTLAHLISILGLPWSHEFQVOMLDMYHRVDRIESMAGEHVNMFADPQ	595	
DB	522	PYISCEYMHAMNGSLGGMKYTELE-----QKYPMTQ-----GCFIMDYID--	562	
QY	596	TNGLIIRVD-----GNK-----KGVFTRDRKPKAAASLRARMTSI	631	
DB	563	--QALLKORYKKEYAYAGDGFDRPTDYSFGANGIVADRKPSPKMGQVVKLYONI	617	

Search completed: March 18, 2005, 23:53.04  
Job time : 46 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2005, 23:42:09 ; Search time 60 Seconds  
(without alignments)  
3492.718 Million cell updates/sec

Title: US-10-757-093-4  
Perfect score: 3354  
Sequence: 1 MKEFLGLSLSLAAPSIGTP.....RKPKAAHSLRARMTSIDKN 634

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1401741 seqs, 330541175 residues  
Total number of hits satisfying chosen parameters: 1401741

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/2/pubppa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubppa/US06\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubppa/US06\_PUBCOMB.pep:\*  
4: /cgn2\_6/ptodata/2/pubppa/US07\_NEW\_PUB.pep:\*  
5: /cgn2\_6/ptodata/2/pubppa/US07\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/2/pubppa/US08\_NEW\_PUB.pep:\*  
7: /cgn2\_6/ptodata/2/pubppa/US08\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/2/pubppa/US09\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubppa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubppa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubppa/US09\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubppa/US09\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/2/pubppa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubppa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubppa/US10\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubppa/US10\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/2/pubppa/US10\_PUBCOMB.pep:\*  
18: /cgn2\_6/ptodata/2/pubppa/US11\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/2/pubppa/US60\_NEW\_PUB.pep:\*  
20: /cgn2\_6/ptodata/2/pubppa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1714	51.1	618	14	US-10-356-088-28
2	1714	51.1	618	16	US-10-799-326-28
3	1714	51.1	1010	6	US-09-118-276-12
4	1714	51.1	1010	16	US-10-705-197A-12
5	1711	51.0	603	14	US-10-161-403-106
6	1711	51.0	603	14	US-10-120-145-6
7	1711	51.0	603	15	US-10-364-649-17
8	1711	51.0	603	15	US-10-364-649-23
9	1711	51.0	603	15	US-10-369-493-842
10	1706	50.9	604	10	US-09-893-525-37
11	1706	50.9	659	10	US-09-893-525-40
12	1706	50.9	850	10	US-09-893-525-42
13	1665.5	49.7	607	15	US-10-369-493-15291

14	1665	49.6	711	14	US-10-338-411-23	Sequence 23, Appl
15	1665	49.6	711	15	US-10-389-640-23	Sequence 23, Appl
16	1664.5	49.6	602	14	US-10-195-518-5	Sequence 5, Appl
17	1439	42.9	607	14	US-10-120-145-8	Sequence 8, Appl
18	1439	42.9	607	15	US-10-364-649-18	Sequence 18, Appl
19	1439	42.9	615	15	US-10-364-649-28	Sequence 28, Appl
20	1439	42.9	618	14	US-10-120-145-4	Sequence 4, Appl
21	1437	42.8	602	14	US-10-120-145-2	Sequence 2, Appl
22	1437	42.8	602	15	US-10-364-649-2	Sequence 2, Appl
23	1437	42.8	602	15	US-10-364-649-8	Sequence 8, Appl
24	1437	42.8	602	15	US-10-364-649-15	Sequence 15, Appl
25	1338	39.9	613	14	US-10-120-145-5	Sequence 5, Appl
26	1338	39.9	613	15	US-10-364-649-16	Sequence 16, Appl
27	1338	39.9	651	15	US-10-421-175-2	Sequence 2, Appl
28	1338	39.9	722	14	US-10-136-841-6	Sequence 6, Appl
29	1338	39.9	722	15	US-10-272-531A-6	Sequence 6, Appl
30	1338	39.9	722	15	US-10-272-483A-6	Sequence 6, Appl
31	1223	36.5	450	14	US-10-421-175-4	Sequence 4, Appl
32	1223	36.5	450	14	US-10-356-088-40	Sequence 40, Appl
33	1223	36.5	450	16	US-10-799-326-40	Sequence 40, Appl
34	1077.5	32.1	598	10	US-09-862-660-2	Sequence 2, Appl
35	1077.5	32.1	598	15	US-10-673-935-2	Sequence 2, Appl
36	993.5	29.6	376	15	US-10-364-649-4	Sequence 4, Appl
37	993.5	29.6	376	15	US-10-364-649-19	Sequence 19, Appl
38	988	29.5	372	15	US-10-364-649-3	Sequence 3, Appl
39	988	29.5	372	15	US-10-364-649-22	Sequence 22, Appl
40	883	26.3	563	15	US-10-364-649-6	Sequence 6, Appl
41	883	26.3	563	15	US-10-364-649-21	Sequence 21, Appl
42	882	26.3	563	15	US-10-369-493-2995	Sequence 2995, Ap
43	756.5	22.6	540	15	US-10-364-649-5	Sequence 5, Appl
44	741.5	22.1	408	15	US-10-369-493-13311	Sequence 13311, A
45	738	22.0	535	15	US-10-364-649-20	Sequence 20, Appl

## ALIGNMENTS

RESULT 1  
US-10-356-088-28  
Sequence 28, Application US/10356088  
Publication No. US2003016753A1  
GENERAL INFORMATION:  
APPLICANT: Yadvav, Jiantun Gene  
TITLE OF INVENTION: Intein-Mediated Protein Splicing  
FILE REFERENCE: C11806  
US NA  
CURRENT APPLICATION NUMBER: US/10/356,088  
CURRENT FILING DATE: 2003-01-31  
PRIOR APPLICATION NUMBER: US 60/354395  
PRIOR FILING DATE: 2002-02-04  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 28  
LENGTH: 618  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Modified GUS protein, with 6x His tags on the N-terminus and  
OTHER INFORMATION: C-terminus  
FEATURE:  
NAME/KEY: PEPTIDE  
LOCATION: (1)..(618)  
OTHER INFORMATION: Modified GUS protein, with 6x His tags on the N-terminus and  
OTHER INFORMATION: C-terminus  
US-10-356-088-28

Query Match 51.1% Score 1714; DB 14; Length 618;  
Best Local Similarity 55.2% Pred. No. 1.2e-136;  
Matches 331; Conservative 92; Mismatches 161; Indels 16; Gaps 9;  
40 VREPORTSRELVDIGIMKFL--ASGLNDTAQPTAPLPKGLBCPVASYNDFISRE 96  
||| |::| ||||| |::| : : : : : ||| |

```

Db 9 VRPEETPREIKKLDGLMAFSLDRENGCIDQ--RWMESALQESRALAVPGSFNDQFADAD 66
QY 97 IHHVGVVYQREVIYVPGKMSQERYLVRAESATTHGRIVYNNRLVAEHVGGYTPFEADYT 156
Db 67 IRNVAGVWVYQREVFIPKMGAGRIYLRFDATVHYGKVVWVNNQEVMEHOGGYTPFEADYT 126
QY 157 ELVAPGEKRLITGVNNELTHETIPPGKITTGNAATGRIOYQHDYFNAGLARSIMLYS 216
Db 127 PYVIAKSVRIYVCNNELMQWITIPPGMIT--DENGKKOSYFHPFNAGHRSVLMYLT 185
QY 217 VPQOHDITVTVTV--DGDNGILNVEVEVANOQTGOIQISVIDEGALVAKASAGQTV 274
Db 186 TPTWVDDITVTVTHVADQCNHASVDQV--VAN--GDVVELRDADQOVVATGQTSGLT 241
QY 275 TIPSVKLMOPGAAYVLOQVNIYVSSGDVDTYNLATGVTVYVAGSQPLINHKPFYFTG 334
Db 242 QVNPPLMQPGEGYLYELCV--TAKSQTECDIYPLRGIRSAVVKGEQFLINHKPFYFTG 299
QY 335 FGKHEDTAVRGKGDPAVYVHDFOLMKWIGANSFRTSHYPYAEVMDPADRNGIYVIDET 394
Db 300 FGHEDADLRGKGFVNVHDLMDWIGANSYRTSHYPYAEVMDPADRNGIYVIDET 359
QY 395 PAVGNIAL-MGVSESGAP-QTTPPAINDKTOEAHKOALRELIARDKXHASVVMMSIAN 452
Db 360 AAVGFNLISIGIFEGAGNKPKEYSEAVNGETOQAHLOAIKELIARDKXHASVVMMSIAN 419
QY 453 EPASHEDGAREYEPILTNLTROLDPTPTFANVGTATYQDLRISDLFVSCINRYFGWY 512
Db 420 EPTROGAREYFAPLAETARKLDPTPTFCVNVMECDANTDTISLDFVLCLNRYFGWY 479
QY 513 SQTGDEEAEALKEIKHGWOKFHRPIVMTYEGADTLAGLSILGLPWESEFQVOMLDM 572
Db 480 VQSGDLEAKVLEKELIAMQEKHOPITITTEYGVDTLAGLSMTYDMWSESEYQCAWLDW 539
QY 573 YHRVFDRIEMAGEHVNFPADFQTNIGIRVDGNKKGVFTDRKPKXAAHSLPARWTSID 632
Db 540 YHRVFDRIEMAGEHVNFPADFQTNIGIRVDGNKKGVFTDRKPKXAAHSLPARWTSID 599

RESULT 2
US-10-799-326-28
; Sequence 28, Application US/10799326
; Publication No. US2004012688A1
; GENERAL INFORMATION:
; APPLICANT: E.I. DuPont de Nemours, & Company
; APPLICANT: Yadav, Narendra S.
; APPLICANT: Yang, Jianjun Gene
; TITLE OF INVENTION: Intein-Mediated Protein Splicing
; FILE REFERENCE: CL1806 US CIP
; CURRENT APPLICATION NUMBER: US/10/799,326
; PRIOR FILING DATE: 2004-03-12
; PRIOR APPLICATION NUMBER: US 60/354395
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 28
; LENGTH: 618
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Modified GUS protein, with 6x His tags on the N-terminus and
; OTHER INFORMATION: C-terminus
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(618)
; OTHER INFORMATION: Modified GUS protein, with 6x His tags on the N-terminus and
; OTHER INFORMATION: C-terminus
US-10-799-326-28

Query Match 51.1%; Score 1714; DB 16; Length 618;
Best Local Similarity 55.2%; Pred. No. 1.2e-136;
Matches 331; Conservative 92; Mismatches 161; Indels 16; Gaps 9;

```

```

QY 40 VRPQRTSREIVNLDGLMKFAL--ASGLNDTAPQWTAPLPKGLCECPVSPAYNDIFISRE 96
Db 9 VRPEETPREIKKLDGLMAFSLDRENGCIDQ--RWMESALQESRALAVPGSFNDQFADAD 66
QY 97 IHHVGVVYQREVIYVPGKMSQERYLVRAESATTHGRIVYNNRLVAEHVGGYTPFEADYT 156
Db 67 IRNVAGVWVYQREVFIPKMGAGRIYLRFDATVHYGKVVWVNNQEVMEHOGGYTPFEADYT 126
QY 157 ELVAPGEKRLITGVNNELTHETIPPGKITTGNAATGRIOYQHDYFNAGLARSIMLYS 216
Db 127 PYVIAKSVRIYVCNNELMQWITIPPGMIT--DENGKKOSYFHPFNAGHRSVLMYLT 185
QY 217 VPQOHDITVTVTV--DGDNGILNVEVEVANOQTGOIQISVIDEGALVAKASAGQTV 274
Db 186 TPTWVDDITVTVTHVADQCNHASVDQV--VAN--GDVVELRDADQOVVATGQTSGLT 241
QY 275 TIPSVKLMOPGAAYVLOQVNIYVSSGDVDTYNLATGVTVYVAGSQPLINHKPFYFTG 334
Db 242 QVNPPLMQPGEGYLYELCV--TAKSQTECDIYPLRGIRSAVVKGEQFLINHKPFYFTG 299
QY 335 FGKHEDTAVRGKGDPAVYVHDFOLMKWIGANSFRTSHYPYAEVMDPADRNGIYVIDET 394
Db 300 FGHEDADLRGKGFVNVHDLMDWIGANSYRTSHYPYAEVMDPADRNGIYVIDET 359
QY 395 PAVGNIAL-MGVSESGAP-QTTPPAINDKTOEAHKOALRELIARDKXHASVVMMSIAN 452
Db 360 AAVGFNLISIGIFEGAGNKPKEYSEAVNGETOQAHLOAIKELIARDKXHASVVMMSIAN 419
QY 453 EPASHEDGAREYEPILTNLTROLDPTPTFANVGTATYQDLRISDLFVSCINRYFGWY 512
Db 420 EPTROGAREYFAPLAETARKLDPTPTFCVNVMECDANTDTISLDFVLCLNRYFGWY 479
QY 513 SQTGDEEAEALKEIKHGWOKFHRPIVMTYEGADTLAGLSILGLPWESEFQVOMLDM 572
Db 480 VQSGDLEAKVLEKELIAMQEKHOPITITTEYGVDTLAGLSMTYDMWSESEYQCAWLDW 539
QY 573 YHRVFDRIEMAGEHVNFPADFQTNIGIRVDGNKKGVFTDRKPKXAAHSLPARWTSID 632
Db 540 YHRVFDRIEMAGEHVNFPADFQTNIGIRVDGNKKGVFTDRKPKXAAHSLPARWTSID 599

RESULT 3
US-09-118-276-12
; Sequence 12, Application US/09118276
; Patent No. US20010011381A1
; GENERAL INFORMATION:
; APPLICANT: BABYCHUK, EBENA;
; APPLICANT: KUSHNIR, SERGEI;
; APPLICANT: DE BLOCK, MARC;
; APPLICANT: INZE, DIRK
; TITLE OF INVENTION: METHODS AND MEANS TO MODULATE PROGRAMMED
; TITLE OF INVENTION: CELL DEATH IN EUKARYOTIC CELLS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SIXBEY, FRIEDMAN, LEEDOM, & PERGUSON
; STREET: 8180 GREENSBORO DRIVE, SUITE 800
; CITY: MCLEAN,
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3-1/2" DISKETTE
; COMPUTER: IBM-COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/118,276
; FILING DATE: 17-JUL-1998
; PRIOR APPLICATION DATA: NONE
; ATTORNEY/AGENT INFORMATION:
; NAME: SCHULMAN, ROBERT M.; SALEN, KENNETH H.
; REGISTRATION NUMBER: 31,196; 43,077
; REFERENCE/DOCKET NUMBER: 6201-0003

```



TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 790-9110  
 TELEFAX: (703) 883-0370  
 INFORMATION FOR SEQ ID NO: 12:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1010 RESIDUES  
 TYPE: AMINO ACID  
 STRANDEDNESS: SINGLE  
 TOPOLOGY: LINEAR

US-09-118-276-12

Query Match 51.1%; Score 1714; DB 9; Length 1010;  
 Best Local Similarity 55.2%; Pred. No. 2,5e-136;  
 Matches 331; Conservative 92; Mismatches 161; Indels 16; Gaps 9;

```

QY 40 VRPQRTSSRELVDLGMKFL--ASGLNDTAQPTAPLPKGLCEPPVASYNDIFISRE 96
DB 409 VRPEPTREIKKLDGLMAFSLDRENGCIDQ--RMWESALQESRAIAVPGSFNDQFADAD 466
QY 97 IDHVGWVYVYQREVIYPKGMSOERYLVRAESATTHGRIVNNRLVAEHVGYTPFEADYT 156
DB 467 IRNYAGNWWYQREVFIPKMGAGRIYLRPAVTHYGVKWNQEVNHEGQGYTFEADYT 526
QY 157 ELVAPGEKRLITGVNNELTHETIPPGKITTGNAQTKRIQTYOHDFYNYAGLARSIMLYS 216
DB 527 PYVIAGKSRYITVCNNELMQITIPPGMYIT--DENGKKQSYHDFPNYAGIHRSMVLYT 585
QY 217 VPQOHIDITVYTDV--DGDNGILNVEVAVANQTTQIQISVIDEDGAIYAKASGAQGY 274
DB 586 TPRTWVDDITVYTHVADCNHASVDWQV--VAN--GDVSEVLADADQOVVATGQTSGLT 641
QY 275 TIPSVLNMGPGAAYLYOLOVNIYVSSGDVVDYTNLATGYRTYKVASGQPLINKPFFYTG 334
DB 642 QVNVPHLMQPGSEYVLELCV--TAKSQTECDIYPLKVGIRSVAVKGEQFLINKPFFYTG 699
QY 335 FGKHEDTAARGKGDPAVWVHDFOLMKWIGANSFRTSHYPYAEVWDFADRNQIYIDET 394
DB 700 FGRHEDADLRGKGFNDVNLVHDAIMDWIGANSYRSHYPYAEMLDMADDEHGIYIDET 759
QY 395 PAVGNIAL--MGVSESGAP--QFTTPDAINDKTOEAKHQAIRELIADKXHASVVMMSIAN 452
DB 760 AAVGFNLISGIGPEAGNKPELYSEAVNGETOQAHLQAIKELIADKXHASVVMMSIAN 819
QY 453 EPASHEGAREYEPETNLTRQLDPTRPITFPAVNGATATYQDRISLDFVSCINRYFGWY 512
DB 820 EPTTRQGARREYFAPLAETRKLDPTRPITCVNVMFCDAHTDTISLDFVCLNRYFGWY 879
QY 513 SQTGLEEAEALAEKELHGMQEKFRHPRIYVWTEYGADTLAHLISILGLPMSSEFOVQMLDM 572
DB 880 VQSGDLETAKEVLEKELHGMQEKHQPITITEYGVDTLAHLISMTYDMSESEYQCAMLDM 939
QY 573 YHRVFPRISSMAGEHYWNPADFOTNLGILRVGNKKGVFTTRDKKPYAAHSLPARWTSID 632
DB 940 YHRVFPRISSMAGEHYWNPADFOTNLGILRVGNKKGVFTTRDKKPYAAHSLPARWTSID 999

RESULT 4
US-10-705-197A-12
; Sequence 12, Application US/10705197A
; Publication No. US20040128704A1
; GENERAL INFORMATION:
; APPLICANT: Elena, Babychuk
; APPLICANT: Sergei, Kuchnik
; APPLICANT: Marc, De Block
; TITLE OF INVENTION: Methods and means to modulate programmed cell death in eukaryotic
; FILE REFERENCE: 58764.0000039
; CURRENT APPLICATION NUMBER: US/10/705,197A
; PRIOR FILING DATE: 2003-11-12
; PRIOR APPLICATION NUMBER: US 09/118,276
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 12

```

LENGTH: 1010  
 TYPE: PRT  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: fusion protein between APP N-terminal domain and GUS protein  
 US-10-705-197A-12

Query Match 51.1%; Score 1714; DB 16; Length 1010;  
 Best Local Similarity 55.2%; Pred. No. 2,5e-136;  
 Matches 331; Conservative 92; Mismatches 161; Indels 16; Gaps 9;

```

QY 40 VRPQRTSSRELVDLGMKFL--ASGLNDTAQPTAPLPKGLCEPPVASYNDIFISRE 96
DB 409 VRPEPTREIKKLDGLMAFSLDRENGCIDQ--RMWESALQESRAIAVPGSFNDQFADAD 466
QY 97 IDHVGWVYVYQREVIYPKGMSOERYLVRAESATTHGRIVNNRLVAEHVGYTPFEADYT 156
DB 467 IRNYAGNWWYQREVFIPKMGAGRIYLRPAVTHYGVKWNQEVNHEGQGYTFEADYT 526
QY 157 ELVAPGEKRLITGVNNELTHETIPPGKITTGNAQTKRIQTYOHDFYNYAGLARSIMLYS 216
DB 527 PYVIAGKSRYITVCNNELMQITIPPGMYIT--DENGKKQSYHDFPNYAGIHRSMVLYT 585
QY 217 VPQOHIDITVYTDV--DGDNGILNVEVAVANQTTQIQISVIDEDGAIYAKASGAQGY 274
DB 586 TPRTWVDDITVYTHVADCNHASVDWQV--VAN--GDVSEVLADADQOVVATGQTSGLT 641
QY 275 TIPSVLNMGPGAAYLYOLOVNIYVSSGDVVDYTNLATGYRTYKVASGQPLINKPFFYTG 334
DB 642 QVNVPHLMQPGSEYVLELCV--TAKSQTECDIYPLKVGIRSVAVKGEQFLINKPFFYTG 699
QY 335 FGKHEDTAARGKGDPAVWVHDFOLMKWIGANSFRTSHYPYAEVWDFADRNQIYIDET 394
DB 700 FGRHEDADLRGKGFNDVNLVHDAIMDWIGANSYRSHYPYAEMLDMADDEHGIYIDET 759
QY 395 PAVGNIAL--MGVSESGAP--QFTTPDAINDKTOEAKHQAIRELIADKXHASVVMMSIAN 452
DB 760 AAVGFNLISGIGPEAGNKPELYSEAVNGETOQAHLQAIKELIADKXHASVVMMSIAN 819
QY 453 EPASHEGAREYEPETNLTRQLDPTRPITFPAVNGATATYQDRISLDFVSCINRYFGWY 512
DB 820 EPTTRQGARREYFAPLAETRKLDPTRPITCVNVMFCDAHTDTISLDFVCLNRYFGWY 879
QY 513 SQTGLEEAEALAEKELHGMQEKFRHPRIYVWTEYGADTLAHLISILGLPMSSEFOVQMLDM 572
DB 880 VQSGDLETAKEVLEKELHGMQEKHQPITITEYGVDTLAHLISMTYDMSESEYQCAMLDM 939
QY 573 YHRVFPRISSMAGEHYWNPADFOTNLGILRVGNKKGVFTTRDKKPYAAHSLPARWTSID 632
DB 940 YHRVFPRISSMAGEHYWNPADFOTNLGILRVGNKKGVFTTRDKKPYAAHSLPARWTSID 999

RESULT 5
US-10-161-403-106
; Sequence 106, Application US/10161403
; Publication No. US20030119104A1
; GENERAL INFORMATION:
; APPLICANT: Perkins, Edward
; APPLICANT: Perez, Carl
; APPLICANT: Lindendaum, Michael
; APPLICANT: Greene, Amy
; APPLICANT: Leung, Josephine
; APPLICANT: Fleming, Elena
; APPLICANT: Stewart, Sandra
; APPLICANT: Shellard, Joan
; TITLE OF INVENTION: CHROMOSOME-BASED PLATFORMS
; FILE REFERENCE: 24601-420
; CURRENT APPLICATION NUMBER: US/10/161,403
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: 60/294,758
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: 60/366,891
; PRIOR FILING DATE: 2002-03-21

```



ORGANISM: *Escherichia coli*  
US-10-364-649-17

US-10-364-649-17

Query Match	Score	DB	Length
51.0%	1711	15	603

Best Local Similarity 55.0%; Pred.No. 2e-136;  
Matches 330; Conservative 93; Mismatches 161; Indels 16; Gaps 9;

Matches 330; Conservative 93; Mismatches 161; Indels 16; Gaps 9;

Qy	40	VRPQRTSSREELVNDJGLTKKFFAL--ASGLNBTACQWTAPLPFGJECPPAPASYNDFISRE	96
Db	2	LRAVETPRRELKKJDDGLWAFSLDRNCNGIDQ--RWMESALQESRAIANPFSQNNDFADAD	59
Qy	97	IHDHGVWYYOREVIVIPKGSQOERYLVAESATHTHGRITYNNRLVAEHVGCTTPEADV	156
Db	60	IRNAYAGNWYOREVFIIPKGMAGQRIYLFDFDAVTHGKVMVNNQEVHEHQGTTPPEADV	119
Qy	157	ELVAPGEKFRLLTICNNNELTHETTPRGKITTCGNALGKRIQYQHDFYVYAGIABSIWLS	216
Db	120	PYVLAGSKSVRTIVCNNELMNOTTIPRGVITV--DENGKKKQSFHFDFYVYAGIHRVWMLYT	178
Qy	217	VPOOHIDITVWTDV--DGDNGLIAYEVEVANNQTCGQIQISVIDEDGAIVAKASAGQTV	274
Db	179	TPNTWDDITVYTHAODCNHASVDWQY-VAN--GDVSVELRADQOVVATGQSTGTL	234
Qy	275	TIPSVKLMQGAAYLYQLOVNIIVGSSGDVDTYNLATGVRTYKVASQOFLINKEPFYTG	334
Db	235	QVNPVPHLMQPEGYLXELCV--TAKSQTECDITPLRVGIRSAVAVGEOFLINHKPFYTG	292
Qy	335	FGKHEDTAVRKGNDPAVMYHDFOLMKITIGANSFPTSHYPAAEVMDPADRNGIYIDET	394
Db	293	FGRHEDADLRKGFENVTLMVDHMLMDVIGANSYRTSHYPAAEVMDWADENGIVIDET	352
Qy	395	PAYGLNLTNL-NGVSESGAP-QTEPTDANDKTQEAHKAIRLIRDRDKNHASVVMVSIAN	452
Db	353	AAVGFENLSLGIFGEFGNKPKEXISEBAVNGETQOHHQAIKELIRDRKNHBSVVMVSIAN	412
Qy	453	EPASHEDGAREYFEBTLNLTQLQDPTREDTTFANVTATYQLODRISDLFPVSCINRYFGMY	512
Db	413	EPDTRPGAREYFAPLAEATRKRLDPTREITTCVNWFFCOAHDTISDLFPDVLCLNRYGMY	472
Qy	513	SGTGDLEAEALAEKELHGMQEKFRPLVMTETGADTLAGHSILGLPMSEEFQVQMDM	572
Db	473	VQSGGLEFAEAEVYLEKELLMQEKLHQPLIITEYGVDTLAGHSMTYDWMSESEYQCAWIDM	532
Qy	573	YHRVFDRIESMAGHEVNNFADPQTMGLIIRVDGNKKGVFTDRDKRKAANSLRAWTSID	632
Db	533	YHRVFDRAVAVAGEQVWNPADPAISQGLIRVGGNKKGIIFTRDRKKSAAFLQKRWTKGN	592
RESULT 8			
US-10-364-649-23			
Sequence 23, Application US/10364649			
Publication No. US20030229921A1			
GENERAL INFORMATION:			
APPLICANT: Richard A. Jefferson and Jorge E. Mayer			
TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE			
TITLE OF INVENTION: PRODUCTS, AND USES THEREOF			
FILE REFERENCE: 190106.405CI			
CURRENT FILING DATE: US/10/364.649			
CURRENT FILING DATE: 2003-02-12			
PRIOR APPLICATION NUMBER: 10/364,649			
PRIOR FILING DATE: 2003-02-12			
PRIOR APPLICATION NUMBER: US 09/270,957			
PRIOR FILING DATE: 1999-03-17			
NUMBER OF SEQ ID NOS: 112			
SOFTWARE: FastSeq for Windows Version 4.0			
SEQ ID NO 23			
LENGTH: 603			
TYPE: PRT			
ORGANISM: Escherichia coli			
US-10-364-649-23			

Matches 330; Conservative 93; Mismatches 161; Indels 16; Gaps 9;

[illegible]

Query Match	51.0%;	Score 1711;	DB 15;	Length 603;
Best Local Similarity	55.0%;	Pred. No. 2e-136;		

```
QY      40 VRPQRTSSRELUNLDGLMKFAL---ASSINDTAQPWTAPLPKGLCECPVASYNDFISRE 96
```

Db	2	LRPVETPTREIKKLGDGMARSLDRENGCIDQ--RMMESALQDSRAIAVPSGFNNQFADAD	59
Qy	97	IHDHGVNVIYQREVIYIPKMSQERLYVRAESATHGRIYVNNRLVAEEVGGITPEADVT	156
Db	60	IRNAGVNWVQREVFIPKMGAGRIYLRFDATVTHGKVVNNQNEWMEHGQGYTFPEADVT	119
Qy	157	ELVAPGEKFLITIGVNNVELTHEITIPGKLTITGNAGKRIQYQHDFFNVYAGLAISIMLYS	216
Qy	120	PYVLAGSVATITCVNNELMQITIPGMVIT--DENKKKQSTFHFDFNYAGIHRSVMLYT	178
Db	217	VPOQHODITVWTDV--DGDNGILINEVEVANQTTGQIQISVIDEDGAIYVAKASGAQTV	274
Qy	179	TPNTWVDITVTVTHVADQCNHASVDWQV-VAN--GDVSVELRDADQQVATGGTSGTL	23
Db	275	TIPSVKLMQOGAAIYIYQLOVNIYVSSGDDVDTYMLATGRTYKVAAGSDFLINGRPFYFTG	33
Qy	235	QVNPFLIMQGESELYELCV--TAKSQTECDITPIPLRVGIRSVAYVKGDFLLNHKRFYFTG	292
Db	335	FGKHEDTAVVNGKGDHDAVYVWHDFOFKWIGANSFRTSHPYAEEVWDPADENGIVVIDET	394
Qy	293	FGRHEDDLRGKGFDMVLVWHDHALMDWIGANSYTSIHPYAEEMLDMDADSHGIVVIDET	353
Qy	395	PVGLNLT--MGVSESGAP--QFTTPDAINDKTOEAKHQAIRELIARDKNHASVYVMSIAN	452
Db	353	AAVGFNLSLIGCEAGNKPELSEEAVNGETQOAHLOAIKELIARDKNHPSVYVMSIAN	412
Qy	453	EPASHDEGAEEYEPPLTNLRQUDPRPTTFPANNGTATAYQDRIISLDFVSCIRYRGWY	512
Db	413	EPDRPGAAEYVPLPAEAIRKIDPRPTTCVNVWCDATDTTISLDFVCLINRYRGWY	472
Qy	513	SQTGDLSEAEALAEKELHGOEKFRHPIYMTVEYGAADTLAHLISLGLPWSEEPQVQMLDM	572
Db	473	VQSGDLETAEKVLEKELLMQOEKRLHQPIITTEYGVDTLAGHLSWYTDWMSSEYGCAMLDM	532
Qy	573	YHVFPRISMAGEHWNVPADPQTNLIGIRVDGNKKGVFTDRKRYKAAASLRARMTSID	632
Db	533	YHVFPRVSAVAVGEQVWNPADFATISQILIRVGANKKGIFTDRKRPKSAALFLQKRMWGMN	592
RESULT 10			
US-09-893-525-37			
Sequence 37, Application US/09893525			
Publication No. US2003012631A1			
GENERAL INFORMATION:			
APPLICANT: Moloney, Maurice M.			
TITLE OF INVENTION: Preparation of Heterologous Proteins on Oil Bodies			
FILE REFERENCE: 9369-172			
CURRENT APPLICATION NUMBER: US/09/893,525			
CURRENT FILING DATE: 2001-06-29			
PRIOR APPLICATION NUMBER: US 09/210,843			
PRIOR FILING DATE: 1998-12-15			
PRIOR APPLICATION NUMBER: US 08/846,021			
PRIOR FILING DATE: 1997-04-25			
PRIOR APPLICATION NUMBER: US 08/366,783			
PRIOR FILING DATE: 1994-12-30			
PRIOR APPLICATION NUMBER: US 08/142,418			
PRIOR FILING DATE: 1993-11-16			
PRIOR APPLICATION NUMBER: US 07/559,835			
PRIOR FILING DATE: 1991-02-22			
NUMBER OF SEQ ID NOS: 42			
SOFTWARE: PatentIn version 3.1			
SEQ ID NO 37			
LENGTH: 604			
TYPE: PR1			
ORGANISM: Artificial Sequence			
FEATURE:			
OTHER INFORMATION: Phase-GUS-phase			
US-09-893-525-37			

Query Match	50.9%;	Score 1706;	DB 10;	Length 604;
Best Local Similarity	54.8%;	Pred. No. 5.5e-136;		
Matches 329;	Conservative 94;	Mismatches 161;	Indels 16;	Gaps 9

Qy	40	VRPQRPSSRRLVNLDCGLMFEAL---ASGLNDPAQSWTBLPFGJACPPAPAYNLIFSR	96
Db	3	LRVELEPTREIKKLDDGLMAFSLDRENCIGDQ--RWMSALQESRAIAPVPSFNQFADAD	60
Qy	97	IHDVGVWVYQOREVIVPKWQSOERYLVRAESATIHGRITVNNRLVAEHVGGYTFPEADV	156
Db	61	IRVAVGVWVYQOREVFIKMGAGORITLVLRDAVTHGKVMWVNNQGEHGGYTFPEADV	120
Qy	157	ELVAPGEKRLITGVNNELTHETTPGKITTGNATGKRIQTYQNDFTVYAGLAISIMLYS	216
Db	121	PVYIAKSKSVRITVCVNNELMTQITPEGWIT--DENGKKKQSYFHDFFNYAGIHSVMLYT	178
Qy	217	VPPQHIQDITVTVDY--DGDNGLIYVEVYVANNQTTGQIQISVIDEDGALVKKASGAGTV	276
Db	180	TEPTWVDDITVTHVHQDCNHSVWQY-VAN--GDVSVELRDADQGVAVTGGGTGTL	233
Qy	275	TIPSVKLMPQGAAYLYQLQVNIIVGSSGVVDVYNLATVTRIVTVKASQGLNKRPFYTG	334
Db	236	QVNVPHLMQPGSGYLTELCV--TAKSQTBODITPLRVGIRSVVAKGQGFLLNHKRPFTG	299
Qy	335	FGHEBDTAVRGKGHDPAYVHDFQLMKNIIGANSFPTSHYPAAEVMYDADRNGIVTDET	394
Db	294	FGHEBDADIRGKGFVDVLMVHDHALMDWIGANSYRTSHYPAAEMLDMADHGGIVTDET	353
Qy	395	PARGNLIAL--MGVSSSGAP--QTFETDAILNDKTOEHKKAITELLARPDNHSVVMWSIAN	455
Db	354	AAAGFSLSIGIEAGNKKPELYSEAVANGETQQAHOAIKELLARPDNHSVVMWSIEIAN	413
Qy	453	EPASHEBDGAREFEPTNLTRQLDPTREPTIFEVANVTATYQADRIISDLFDVSCIRRYGMY	512
Db	414	EPDTRQGARREVFAPLAETATKRLDPTTRITCVNWFPCDANHDTISDLFDVLCNRRYGMV	473
Qy	513	SQTGDLLEBAALBKELHGWQKQFHRPIVMTYEGADTLAHLSTLGLPWSEEFQVQMLDM	572
Db	474	VQSGDLETEAKVLEKELLAMQEKHQPIITIEYGVDTLAGHSYWTMWSSEYQCAWILDM	533
Qy	573	YHHPVPRISMGHEVYVNNADPQTNLGLIRVDGNKKGVPTPDRKPKAAHSLRARWMSID	633
Db	534	YHHPVPRVAVVGEVQVNNADPATSYQGLIRVGNKKGLPTDRKPKSAALFLQGRWGMN	593

```

RESULT 11
US-09-893-525-40
; Sequence 40, Application US/09893525
; Publication No. US20030126631A1
; GENERAL INFORMATION:
; APPLICANT: Moloney, Maurice M.
; APPLICANT: Van Rooijen, Gifs
; TITLE OF INVENTION: Preparation of Heteroc
; FILE REFERENCE: 9369-172
; CURRENT APPLICATION NUMBER: US/09/893,525
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 09/210,843
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: US 08/846,021
; PRIOR FILING DATE: 1997-04-25
; PRIOR APPLICATION NUMBER: US 08/366,783
; PRIOR FILING DATE: 1994-12-30
; PRIOR APPLICATION NUMBER: US 08/142,418
; PRIOR FILING DATE: 1993-11-16
; PRIOR APPLICATION NUMBER: US 07/659,835
; PRIOR FILING DATE: 1991-02-22
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1.
; SEQ ID NO 40
; LENGTH: 659
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: phas-oleo GUS-phas
US-09-893-525-40

```

Query Match	50.9%	Score 1706	DB 10	Length 659
Best Local Similarity	54.8%	Pred. No. 6,2e-136		
Matches 329	Conservative 94	Mismatches 161	Indels 16	Gaps 9
QY	40	VRPORTSRELEVINLDTGKMFAL---ASGLINDAPQWTPAPLPGLGECPPASYNDFISRE	96	
DB	58	LRPVETPRELEIKLUDGLMAFSLDRNCGDQ--RWMESLQSRALAYVGSNDNDQPADMD	115	
QY	97	IHDHVGWYVYQREVIVIPKGSQERYLVAESAETHGRILYVNNRLVAEHVGYTTPEDAVT	156	
DB	116	IRNAYAGWVYQREVEIPKGMAGORIVLRFDAVTHYGKVVNNQVEWHEOGGYTPEADVT	175	
QY	157	ELVAPGEFRFLTICGNMELTHETTPRKITITGNALNGKRIQTYQHDPYNYAGLABISIMLS	216	
DB	176	PVYLAGSKSVRTTCVNNELMOTLTPPGWVT--DENGKKQSYFHOFENYAGIHRSMVLT	234	
QY	217	VPOOHIDITVTVTDV--DGDNGLIYVEYVANQTTGOIQLSIDDEGALVAASGAQGTV	274	
DB	235	TPNFWVDITVYTHVAQDCNHSADVWQV-VAN--GDVSVELRDADQOVVATGQSTGTL	290	
QY	275	TIPSVKLMQPGAAVLYLOQVNIIVGSSGDVVDVTYNLATGVRIVKVASGQFLNGKPPFTG	334	
DB	291	QVNVPHLMQPEBGVLYELCV--TAKSQTECDIYPLRVGIRSAVAVGQQLIHNKQPFYFG	348	
QY	335	FGKHEDTAVRKGKHPAYMADPQLMKITIGANSFPTSHYPAVEEWMDFADRGIYVIDET	394	
DB	349	FGREHEDDLRKGKGFNPVLMVHDHALMDMIGANSYRTSHPYAEEMLDMADDEHGIYVIDET	408	
QY	395	PAYGLNIAL--NGVSESGAP--QTFPTDAILNDKIQEAKHQAIRLIRADKNHAAVMMSTAN	452	
DB	409	AAVGFSLSLGIGFEGKPKKPELYSEAVNGETQOHLQIKLIRADKNHPSVMMSTAN	468	
QY	453	EPASHEDGAREYFEBLTNLTRQOLDPRTDITTFANVGTATYQOLDRIISDLFVDSCLNRYFGWY	512	
DB	469	EPDTRPGARERYPAVLAEATRLDPTRPITTCVNWVFCOAHNTDTISDLFVULCLNRYGYW	528	
QY	513	SGTGDLEAEALAEKELHGMQEKFRPIYMTETYGADTLAGLHSITGLPMSEEFQVQMDM	572	
DB	529	VQSGDLEAEAEVYLEKELELMAOEKHLQPIITIEYGVDTTLAGLHSMTDMMSEERYQCAMLDM	588	
QY	573	YHAFVDRISMAGEHVNNFADPQTNLGIIRVDGNKKGVPTPRKPKRAAHSIRARWTSID	632	
DB	589	YHAFVDRISAVVGEQVWNPADFATISQGLIRVGANKKGIPTDRKPKRAAFLIQKMTGN	648	

OTHER INFORMATION: phae-caleo-GUS-phae

US-09-893-525-42

Query Match 50.9%; Score 1706; DB 10; Length 850;  
Best Local Similarity 54.8%; Pred. No. 9, 1e-136;  
Matches 329; Conservative 94; Mismatches 161; Indels 16; Gaps 9;

40 VRPGRSSRELVNLDGIMKFPAL--ASGLNDIAQPTAPLPGLEGEPYPASINDIFISRE 96  
Db 249 LRPEVETREIKKLIDGIMAFSLDRENCGIDQ--RWMSALQESRAIIVAGSINDPFDAD 306  
97 IHDHVGWVQREEVIPKGSQERYVRAASATTHGRIVVNRVLAEHGVGTTPREADVT 156  
Db 307 IKNAGWVWQREEVIPKMGAGQIVLRFDATVTHGKVVANNQSEWHEOGGTTPEADVT 366  
157 ELVAPGEKFLITGVNNELTHETIPPKGTTGNATGKRIGTYQHDFFYNYGLARSIMYS 216  
Db 367 PYVLAGSVAITVCNNELNMWOTIPPGWIT-DEKKKKQSIFHDPFNVAGIHRSMVLYT 425  
217 VPQOQHODITVTVTDV--DGDNGILNVEVANNQTTGQIOISVIDBDGATVAKASAGQTV 274  
Db 426 TPTWTWDDITVTVTHVADQCNHASVDMQV-VAN--GDVSVELRLDADQGVVATGCGTGT 481  
275 TTSVTLMOGGAAYVLQLOQNIYVSSGSDVDVRYNLATGRTKVGASQFLINGKPFYFG 334  
Db 482 QVNPVHLMOGEGYLYELCV-TAKSQTECDIPLRLVGRISVAIVGQQCLINHKPFYFG 539  
335 FGKHEDTAVVGGKHDDAYVWVHDFOLMKWIGANSFRTSHPYAEVWDPFADRNGLVVIDET 394  
Db 540 FGRHEDADLRGKFDVNLVWHDALMDWIGANSYRSHPYAEMLDMADDEHGIIVIDET 599  
395 PAVGLNIAL-MGVSESGAP-QTFTPDAINDKTOEAAKQAIIRELIARDKXHASVVMWSTAN 452  
Db 600 AAVGFSISLIGIEAGNKKPKELYSEEAENVGETQOAHLLQAIKELIARDKXHPVVMWSTAN 659  
453 EPVSHDEGAEEYEPPLTNLTROLDPRPLTPFNAGVATQQLRISGLDPVSGINRYPGY 512  
Db 660 EPTRRPGAAEYAPPLAEATRKLDPRPLTCVNVWFCDAAHTTISLDFPVLCNLNRYGY 719  
513 SQTGDELEAAALEKEKLHGQEKFRPIVWTEYGAOTTLAHLISLGLPVSSEFQVQMLDM 572  
Db 720 VQSGDLETAKEVLEKELMLAQEKLHQPILITTEYGVDTLAGLSMTMDMSEERYQCAWMD 779  
573 YHRVFDRIESMAGEHVNFPADFQTNIGIIRVDGNKKGVETDRDKPKAAASHLRARWTSID 632  
Db 780 YHRVFDRESVAIVGEQVWNFADFATSGILRVGKNKKGIETDRDKPKSAFLQKSWTGMN 839

RESULT 13  
US-10-369-493-15291  
Sequence 15291, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 15291  
LENGTH: 607  
TYPE: PRT  
ORGANISM: Agrobacterium tumefaciens  
US-10-369-493-15291

49.7%; Score 1665.5; DB 15; Length 607;

Best Local Similarity 53.8%; Pred. No. 1.5e-132;  
Matches 324; Conservative 94; Mismatches 167; Indels 17; Gaps 10;

```

QY 38 IKVRFPTSSRELVLNDGLMKFAL--ASGLNDTAQPMVAPLPGKECPVPSPSYNDIFIS 94
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 LMLRPETPTREIKKLIDGLMAFSLDRENGCIDQ--RMWESALQESRAIAVPSPFNDQFAD 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 95 REIHDVGVWYVYQREIVIPKGMQSEERYLVRAESATHGRIVYNNRLVAEHVGYTPPEAD 154
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 ADIRNAGVWYVYQREIVIPKGMAGQRIIVRPDAVTHYKVMWVNNQGVMEHQGYTPPEAD 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 155 VTELVAPEKEXRLITGVNNELTHETTPPKKITTGATGKRIQTYCHDFNVYGLRSIML 214
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 VTEYVLAGSKSVRLTVCNNELMWTIPPGVIT--DENGKKOSYFPDFNVYAGIHSVWL 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 215 YSPQOHIODITVYTDV--DGDNGLINYEVEVANQTTGOIQLSVIDEDGAIYAKASGAG 272
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 YTPNTWVDITVYTHVADCNHASVDWQY-VAN--GDVVELRADQGVVATGQTSG 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 273 TVTIPSVKLMQPGAAVLYQLQVNIIVSGSDVDVTYNLATGVTYKVASGQFLNGKPYF 332
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 TLQVNVNPHLMQPGGYLYELCV--TAKSQTECDIYPLRVGIRSVAVKGOFLINHKPYF 295
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 333 TGGKHEDTAVRKGDHPATWMDPOLMKWIGANSRPTSHYPAAEVMPFADENGIVYD 392
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 296 TGFGRHEDADLRGKGFNVLMVDHMLDWIGANSYSTSHYPAEEMLDMADEHGLVYD 355
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 393 ETPAVGNIAL-MGVSSGAP-QTFPTDAINDKTOEAKQAIRELIAARDKNHASVVMWSI 450
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 356 ETAAVGNLISGIFEAQNKPKELYSEAVNGETQAHLOAIKELLARDQNHSSVVMWSI 415
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 451 ANEPASHEDGAREYFEPLTNLTROLDPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 510
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 416 ANEDPTPO-VHGNISPLAETRLDPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 474
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 511 MYSTGDLSEEALEKELEHGMOKFRPIVMTVEGADTLGLHSILGLPWSEEFQVOML 570
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 475 WYVOSGDLTETKEKLEKELAMOKELHQPITITTEYGVDTLAGHSMTDMWSEEQACWL 534
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 571 DMVHRVFDRIEBSMGEHVMNFPQTNLGIIRVDGNKKGVFTPRRKRKAANHSIRAWTS 630
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 535 DMVHRVFDRIEBSMGEHVMNFPQTNLGIIRVDGNKKGVFTPRRKRKAANHSIRAWTS 594
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 631 ID 632
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 595 MN 596
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 14  
US-10-338-411-23

```

; Sequence 23, Application US/10338411
; Publication No. US20030153045A1
; GENERAL INFORMATION:
; APPLICANT: Butt, Tauseef
; APPLICANT: Butt, Tauseef
; APPLICANT: Tran, Hiep
; APPLICANT: Weeks, Stephen
; APPLICANT: Malakhova, Oksana
; APPLICANT: Malakhova, Michael
; TITLE OF INVENTION: Methods and Compositions for Protein
; FILE REFERENCE: 1955-2792US1
; CURRENT APPLICATION NUMBER: US/10/338,411
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: 60/346,449
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 23
; LENGTH: 711
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Sequence

```

US-10-338-411-23

Query Match 49.6%; Score 1665; DB 14; Length 711;  
Best Local Similarity 52.4%; Pred. No. 2.2e-132;  
Matches 328; Conservative 98; Mismatches 178; Indels 22; Gaps 12;

```

QY 19 TPAARHPFVEMMDQ--HEQPL----IKVRFPTSSRELVLNDGLMKFAL--ASGLNDTAQ 70
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 85 TPEDLMDENDIILAHNEQIGQEMFMRLPVETPTREIKKLIDGLMAFSLDRENGCIDQ--R 142
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 71 PWTAPLPKGLCECPVPASYNDFISREIHDVGVWYVYQREIVIPKGMQSEERYLVRAESATH 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 143 WMSALQESRAIAVPSPFNDQFADIRNAGVWYVYQREIVIPKGMAGQRIIVRPDAVTH 202
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 131 HGRIVYNNRLVAEHVGYTPPEADVTELVAPEKEXRLITGVNNELTHETTPPKKITTGNA 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 203 YGKVMWVNNQGVMEHQGYTPPEADVTVYVLAGSKSVRLTVCNNELMWTIPPGVIT--DE 261
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 191 TGRKIOTYQHDFFYVYAGIARSIMLYSVPOQHIODITVYTDV--DGDNGLINYEVEVANQ 248
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 262 NGKKQSYFDFENVYAGIHSVWLTPNTWVDITVYTHVADCNHASVDWQY-VAN-- 318
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 249 TGOIQLSVIDEDGAIYAKASGAGTPTTIPSVKLMQPGAAVLYQLQVNIIVSGSDVDVTYN 308
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 319 -GDVVELRADQGVVATGQTSGLQVNVNPHLMQPGGYLYELCV--TAKSQTECDIY 375
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 309 LATGVTYKVASGQFLNGKPYFTGPKHEDTAVRKGDHPATWMDPOLMKWIGANSF 368
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 376 LRVGIRSVAVKGOFLINHKPYFTGPKHEDADLRGKGFNVLMVDHMLDWIGANSY 435
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 369 RTSHPYAEVMPFADENGIVYDPTPAVGNIAL-MGVSSGAP-QTFPTDAINDKTOE 426
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 436 RTSHPYAEVMPFADENGIVYDPTPAVGNIAL-MGVSSGAP-QTFPTDAINDKTOE 495
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 427 AHKQAIRELIAARDKNHASVVMWSIANEPASHEDGAREYFEPLTNLTROLDPTPTPT 486
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 496 AHKQAIRELIAARDKNHASVVMWSIANEPDTPPO-VHGNISPLAETRLDPTPTPTPT 554
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 487 GTATYQDLRIISDLFPVSCINRYFYNSOTGDLSEEALEKELEHGMOKFRPIVMTVEG 546
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 555 MFCDAHTDITISDLFPVLCINRYGYVOSGDLTETKEKLEKELAMOKELHQPITITTEY 614
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 547 ADTLAGHSILGLPWSEEFQVOMLDMVHRVFDRIEBSMGEHVMNFPQTNLGIIRVDGN 606
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 615 VDTLAGHSMTDMWSEEQACWLDMVHRVFDRIEBSMGEHVMNFPQTNLGIIRVGSN 674
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 607 KKGVFTDRKRPRKAAHSIRAWTSID 632
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 675 KKGIFTRDRKRKPSAFLQKRWTKMN 700
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15  
US-10-389-640-23

```

; Sequence 23, Application US/10389640
; Publication No. US20040018591A1
; GENERAL INFORMATION:
; APPLICANT: Butt, Tauseef
; APPLICANT: Butt, Tauseef
; APPLICANT: Tran, Hiep
; APPLICANT: Weeks, Stephen
; APPLICANT: Malakhova, Oksana
; APPLICANT: Malakhova, Michael
; TITLE OF INVENTION: Methods and Compositions for Protein
; FILE REFERENCE: 1955-P02972US2
; CURRENT APPLICATION NUMBER: US/10/389,640
; PRIOR FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: 10/338,411
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: 60/346,449
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 23

```

/ LENGTH: 711  
/ TYPE: PRT  
/ ORGANISM: Artificial Sequence  
/ FEATURE:  
/ OTHER INFORMATION: Synthetic Sequence  
US-10-389-640-23

Query Match 49.6%; Score 1665; DB 15; Length 711;  
Best Local Similarity 52.4%; Pred. No. 2, 2e-132;  
Matches 328; Conservative 98; Mismatches 178; Indels 22; Gaps 12;

QY 19 TPAARHPNEMTQ--HEQPL--IKVPORTSSRELVLNLDGLMKPAL--ASGLNDTAQ 70  
DB 85 TPEIDLMDENDDIIEAREQIGGMEFMLRPVETPTREIKKLDGLMAFSLDRENGCIGD--R 142  
QY 71 PWTAPLPKGLCECPVPSYNDIFISREIHVGVVYQREVIYVPGKWSQERILVRAESATH 130  
DB 143 WMSALQESRAIAVPGSFNDQFADADIRNAGVWYQREVFIPKMGAGORIVLRFDAVTH 202  
QY 131 HGRIVNNRLVAEHVGGYTPFEADVTELVAPGEKFRITGVNNELTHETIPGKITTGNA 190  
DB 203 YGKVMVNNBOEVMHOGYTPFEADVTPYIAGKSVRITVCNNELMWOTIPGGMVIT-DE 261  
QY 191 TGRRIQTYOHDFYVYAGLARSIMLYSVPOQHIDITVVTDV--DGDGLINYEVEVANQT 248  
DB 262 NGKKKQSYFPHDFENVAGIHRSVMLYTTPTWVDITVTVHVAQDCNHASVDMQV--VAN-- 318  
QY 249 TGOIQISVIDEDGAIYAKASGAOGYTTISBVKLMQGAATLYOLOVNIYVSSGDVVDYTN 308  
DB 319 -GDVSYELDADADQVATGQTSGLQVNPMLMQPGEGLYELCV--TAKSQTECDIYP 375  
QY 309 LATGVRTVYVAGSQFLINGKPFYFTGFGKHEDTAVRGKGDPAVYVHDFOIMKMGANSF 368  
DB 376 LRVGIRSVAVKQGFILNHKPFYFTGFGKHEDADLRKGFQDNLVMDHADMIGANSY 435  
QY 369 RFSHYPAEAEVMDPADRNGIIVIDETPAVGLNIAL--MGVSESGAP--QTFTPDAINDKTOE 426  
DB 436 RFSHYPAEAEVMDPADRNGIIVIDETPAVGLNIAL--MGVSESGAP--QTFTPDAINDKTOE 495  
QY 427 AHKQAIRBELIARDKNTASVYVMSIANEPASHEDGAREFEPLTNLTROLDPTPTIFANY 486  
DB 496 AHLQAIKELIARDKNTASVYVMSIANEPASHEDGAREFEPLTNLTROLDPTPTIFANY 554  
QY 487 GATATVQLDRISDLFDVSCINRYFGWYSQGDLEAEAEALEKELHGMQEKFRPIVMTBYG 546  
DB 555 MCDADHTDTISDLFDVCLNRYFGWYVQSGDLEAEAEALEKELHGMQEKLDHPIITTEYG 614  
QY 547 ADTLAHLHSILGLPMSEEFQVQMLDMYHRVYFDRIESMAGEHVNVPADFOINLGIIRVDGN 606  
DB 615 VDTLAGLHMYTDMWSEBEYQCAWLDMYHRVYFDRIVSAVVBQVNNPADFATISQILAVGKN 674  
QY 607 KKGVFTRDRKPKAAHSLPARWTSID 632  
DB 675 KKGIFTRDRKPKSAFLQKRWTKMN 700

Search completed: March 18, 2005, 23:52:14  
Job time : 62 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: March 18, 2005, 23:45:19 ; Search time 43 Seconds  
(without alignments)  
1100.640 Million cell updates/sec

Title: US-10-757-093-4  
Perfect score: 3354  
Sequence: 1 MKFLGLSLSLAAPSIGP.....RKPKAAASLRARWTSIDKN 634

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/iaa/6C\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/iaa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1721.5	51.3	832	3	US-08-630-820-7 Sequence 7, Appl1
2	1721.5	51.3	832	4	US-09-273-453-7 Sequence 7, Appl1
3	1714	51.1	1010	4	US-09-118-726-12 Sequence 12, Appl1
4	1711	51.0	603	3	US-09-149-727-6 Sequence 6, Appl1
5	1711	51.0	603	4	US-09-270-957-17 Sequence 17, Appl1
6	1711	51.0	603	4	US-09-270-957-23 Sequence 23, Appl1
7	1706	50.9	604	4	US-09-893-525-37 Sequence 37, Appl1
8	1706	50.9	659	4	US-09-893-525-40 Sequence 40, Appl1
9	1706	50.9	850	4	US-08-882-704-5 Sequence 42, Appl1
10	1664.5	49.6	602	2	US-08-882-704-5 Sequence 5, Appl1
11	1664.5	49.6	602	4	US-09-151-957-5 Sequence 5, Appl1
12	1664.5	49.6	602	6	5432081-2 Patent No. 5432081
13	1664.5	49.6	602	6	5432081-2 Patent No. 5432081
14	1664.5	49.6	1242	4	US-09-488-270A-2 Patent No. 5268463
15	1638.5	48.9	600	6	5268463-2 Patent No. 5268463
16	1638.5	48.9	600	6	5268463-2 Patent No. 5268463
17	1439	42.9	607	3	US-09-149-727-8 Sequence 8, Appl1
18	1439	42.9	607	4	US-09-270-957-18 Sequence 18, Appl1
19	1439	42.9	615	4	US-09-270-957-28 Sequence 28, Appl1
20	1439	42.9	618	3	US-09-149-727-4 Sequence 4, Appl1
21	1437	42.8	602	3	US-09-149-727-2 Sequence 2, Appl1
22	1437	42.8	602	4	US-09-270-957-2 Sequence 2, Appl1
23	1437	42.8	602	4	US-09-270-957-2 Sequence 8, Appl1
24	1437	42.8	602	4	US-09-270-957-15 Sequence 15, Appl1
25	1338	39.9	613	3	US-09-149-727-5 Sequence 5, Appl1
26	1338	39.9	613	4	US-09-270-957-16 Sequence 16, Appl1
27	1338	39.9	651	4	US-09-715-858-2 Sequence 2, Appl1

28	1323	39.4	648	4	US-09-715-858-4 Sequence 4, Appl1
29	1077.5	32.1	598	4	US-09-862-660-2 Sequence 2, Appl1
30	1027	30.6	500	4	US-09-949-016-11697 Sequence 11697, A
31	993.5	29.6	376	4	US-09-270-957-4 Sequence 4, Appl1
32	993.5	29.6	376	4	US-09-270-957-19 Sequence 19, Appl1
33	988	29.5	372	4	US-09-270-957-3 Sequence 3, Appl1
34	988	29.5	372	4	US-09-270-957-22 Sequence 22, Appl1
35	883	26.3	563	4	US-09-270-957-6 Sequence 6, Appl1
36	883	26.3	563	4	US-09-270-957-21 Sequence 21, Appl1
37	756.5	22.6	540	4	US-09-270-957-5 Sequence 5, Appl1
38	738	22.0	535	4	US-09-270-957-20 Sequence 20, Appl1
39	473.5	14.1	385	4	US-09-270-767-43634 Sequence 43634, A
40	413.5	12.3	1053	4	US-09-134-000C-5361 Sequence 5361, Ap
41	395.5	11.8	282	4	US-09-634-238-345 Sequence 345, App
42	390.5	11.6	1039	4	US-09-501-136-2 Sequence 2, Appl1
43	372.5	11.1	1010	4	US-09-654-448-2 Sequence 2, Appl1
44	372.5	11.1	1010	4	US-09-759-152A-2 Sequence 2, Appl1
45	372.5	11.1	1121	1	US-07-789-915A-2 Sequence 2, Appl1

## ALIGNMENTS

RESULT 1  
US-08-630-820-7  
Sequence 7, Application US/08630820  
Patent No. 6008023  
GENERAL INFORMATION:  
APPLICANT: OPPER, Martin  
APPLICANT: BOSSLET, Klaus  
TITLE OF INVENTION: CYTOPLASMIC EXPRESSION OF ANTIBODIES,  
TITLE OF INVENTION: ANTIBODY FRAGMENTS AND ANTIBODY FRAGMENT FUSION MOLECULES  
NUMBER OF INVENTIONS: 7  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/630,820  
FILING DATE: 10-APR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 19513676.4  
FILING DATE: 11-APR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: GRANADOS, Patricia D.  
REGISTRATION NUMBER: 33,683  
REFERENCE/DOCKET NUMBER: 18748/306  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ. ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 832 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-630-820-7  
Query Match 51.3%; Score 1721.5; DB 3; Length 832;  
Best Local Similarity 52.9%; Pred. No. 2.4e-134;  
Matches 339; Conservative 98; Mismatches 179; Indels 25; Gaps 11;

```

APPLICATION NUMBER: 08/630,820
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: GRANADOS, Patricia D.
REGISTRATION NUMBER: 33,683
REFERENCE/DOCKET NUMBER: 18748/306
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 832 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-273-453-7

Query Match      51.3%; Score 1721.5; DB 4; Length 832;
Best Local Similarity 52.9%; Pred. No. 2.4e-134;
Matches 339; Conservative 98; Mismatch 179; Indels 25; Gaps 11;

Cy      8      SLSLAIPASLTGPA----ARHPREMENTQHGPV-----IKRPORTSREELVNDLGIMK 58
Db      190    SVTVPESSSLGTQTYTTCNVNKHKPSNTKYDKRVELGSGSGSNVRVETPTREIKKLIDGIMA 249
Cy      59      FAL---ASGNDTAQPTAPLPKGLCECPVASYNDFISREIDHVGWVWYQREVIYVKG 115
Db      250    FSLDRNMGIDQ--RWMSALQESRALAVPESFNDQFADADIRVYAGVWVYQREVIYK 307
Cy      116    WSQRRYLVRASATIHGRILYNNRFLVAHVGGYTPFEADVTELYAPEGKFLITIGNNEL 175
Db      308    WAGRIYALRPDAVTHYKGVWVWVNDVEHOGGYTPFEADVTPRYIAGSVAITVCNNEL 367
Cy      176    THEITPGKTTGATGKRIQTYOHDFYNVAGLARSIMLYSPQOHDITVWTDV--DG 233
Db      368    NWQITTPGQWYIT--DENGKKOSYHNHFNNYAGIRSWMLYTPTEWTDITVTVHVAQDC 426
Cy      234    DNGILNIEVEYANOTTQOIQISVIDEDGAIYAKASGAQGITVITPSVKLMQGAALYLQ 293
Db      427    NHAQVDMQV--VAN---GDVSELADDAQOVVATGQTSGLIQVNPPLHMQEGEYLIELC 482
Cy      294    VNIYSGSGDVVDYTNLATGVTFTVYAGSQFLINGKPYFTFGFGKEDTAVYAGKHDPAYM 353
Db      483    V--TAKSQTECDIYPLRIVGIRSVAVKSGQFLINKKPYFTFGFGHEADLKGKGFNVLM 540
Cy      354    VHNDQLMKWIGANSFRTSHYPAEAWDFDNRGIVITDETPAYGLNAL--MGVSESGAP 412
Db      541    VHDALMDWIGANSYRTSHYPAEEMLDMADEHGIIVIDEPAAGFNLSLIGIFEGAKNP 600
Cy      413    -QTFTPAINDKTOEAKHQAIRELIARDKNASVYVMSIANEPASHEDEGAREYEPFLTNL 471
Db      601    KELYSBEAVNGEFOQAHLQAKELIARDKNHPSVYVMSIANEBDTPRGAGREYFAPLAEA 660
Cy      472    TRQLDPRPITFANVGATYQLDRIISDLFDVSCINRYFGWTSQGDLEAEAPALEKEIHG 531
Db      661    TRKLDPRPITCVAVMFCDAHTDITISLFDVLCINRYYGWVQSGDLETAKVLEKELLA 720
Cy      532    WQEKFNRPYVTEEGATTLAGLHSLILGPMSEEEQOVQMLDMYHVFVPRIDEMAGEHWNF 591
Db      721    WQEKLDHPITITTEGVDTTLAGLHSMYTDMMSEEQCAMLDMYHVFVPRVSAVSGEOWNF 780
Cy      592    ADFQTNLGIIRVDGNKKGVFTRDCKPRAAAASHLRARMTSID 632
Db      781    ADFATSGCILRVGSKKGIPTFRDKKPSAAFLQKRTGMN 821

RESULT 3
US-09-118-276-12
Sequence 12, Application US/09118276
Patent No. 6693185
GENERAL INFORMATION:

```

APPLICANT: BABYCHUK, ELENA;  
 APPLICANT: KUSHNIR, SERGEI;  
 APPLICANT: DE BLOCK, MARC;  
 APPLICANT: INZE, DIRK  
 TITLE OF INVENTION: METHODS AND MEANS TO MODULATE PROGRAMMED  
 CELL DEATH IN EUKARYOTIC CELLS  
 NUMBER OF SEQUENCES: 21  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: SIXEY, FRIEDMAN, LEEDOM, & FERGUSON  
 STREET: 8180 GREENSBORO DRIVE, SUITE 800  
 CITY: MCLEAN,  
 STATE: VIRGINIA  
 COUNTRY: USA  
 ZIP: 22102  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3-1/2" DISKETTE  
 COMPUTER: IBM-COMPATIBLE  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: ASCII  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/118,276  
 FILING DATE: 17-JUL-1998  
 PRIOR APPLICATION DATA: NONE  
 ATTORNEY/AGENT INFORMATION:  
 NAME: SCHULMAN, ROBERT M.; SALEN, KENNETH H.  
 REGISTRATION NUMBER: 31,196; 43,077  
 REFERENCE/DOCKET NUMBER: 6201-0003  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 883-9110  
 TELEFAX: (703) 883-0370  
 INFORMATION FOR SEQ ID NO: 12:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1010 RESIDUES  
 TYPE: AMINO ACID  
 STRANDEDNESS: SINGLE  
 TOPOLOGY: LINEAR  
 US-09-118-276-12

Query Match 51.1%; Score 1714; DB 4; Length 1010;  
 Best Local Similarity 55.2%; Pred. No. 1,4e-133;  
 Matches 331; Conservative 92; Mismatches 161; Indels 16; Gaps 9;  
 40 VRPQRTSRELVLNDGLMKPAL--ASGLNDTAQPTAPLPKGLCEPVASYNDIFISRE 96  
 409 VRPVEPTREIKKLKLDGLMAFLDRENCIDQ--RWMESALOESRAIAVPSFNDQFADAD 466  
 97 IHDHVGWVYQREVEIVPKGWSQERLYVRASATHGRIVYNNRLVAEHVGYTPFEADVT 156  
 467 IRNYAGWVWYQREVEIVPKGWSQERLYVRASATHGRIVYNNRLVAEHVGYTPFEADVT 526  
 157 ELVAPGEKRLITGVNNELTHEITIPPKKITGTGATGKRIOTYQHDFFNYAGLARISWLYS 216  
 527 PYVIAKGSVRIITVCNNELMWQITIPGMVIT--DENGKKQSYPHDFPNYAGIHSVWLYT 585  
 217 VPQOHODITVYVDV--DGNGLINYEVEVANQTTGOIOISVIDEDGAIYAKASGAQGV 274  
 586 TPTWVDDITVTVHVAODCNHASVDWQV--VAN--GDVSEYLRDADQOVVATGQSTGTL 641  
 275 TIPSVKLMQFGAAYLYQLQVNIIVSSGDDVDTYNLATGVRTVYVAGSQPLINKPFFYTG 334  
 642 QVNNPHLMQGBEGYIVELCV--TAKSQTECDIPLRGIRSVAVKGBQPLINKPFFYTG 699  
 335 FGHEDTAVRGKHDPAVYVNDPQMKWIGANSFRTSHYVYAEVWDFADRNGLIVTDET 394  
 700 FGHEDADLRGKGFDDVNLVMDHADMWIGANSYRTSHYVYAEVWDFADRNGLIVTDET 759  
 395 PAYGLNIAL-MGVSSEBAP--QTFPTPAINDKTOEAKHQAIRELIARDKNASVVMMSIAN 452  
 760 AAVGFNLISLIGIEAGNKPEKLYSEEVNGETQAHQAIKELIARDKNPSVVMMSIAN 819  
 453 EPASHEDGAREYEPRLNLTROLDPTRPITFANVGATAYQLDRISDLFDVSCINRYFGWY 512  
 820 EPOTRPGAREYFAPRLAEATRKLDPTRPICVNVVFCDAHTDITISDLFDVLCINRYFGWY 879

513 SOTGDLEBAEALAEKELHGWQEKFRHPIVWTEYGADTLAGHSILGLPMSEEFQVOMLDM 572  
 880 VQSGDLETAKEVLEKELHGWQEKHOPITTEGVTTLAGHSIMYDMSEEFQVOMLDM 939  
 573 YHRFPDRISMAGEHVMNFPADFTQNLGIIRVDNKKGVFTDRDKPKRAAHSRLARWTSID 632  
 940 YHRFPDRISAVVGEQVWNFADFATSGILRVGNKKGIPTDRDKPKSAFLQKRWTKMN 999  
 RESULT 4  
 US-09-149-727-6  
 ; Sequence 6, Application US/09149727  
 ; Patent No. 6391547  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Jefferson, Richard A.  
 ; APPLICANT: Kilian, Andrezej  
 ; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND  
 ; FILE REFERENCE: 190106.405  
 ; CURRENT APPLICATION NUMBER: US/09/149,727  
 ; CURRENT FILING DATE: 1998-09-08  
 ; EARLIER APPLICATION NUMBER: US 60/058,263  
 ; EARLIER FILING DATE: 1997-09-09  
 ; NUMBER OF SEQ. ID NOS: 71  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 6  
 ; LENGTH: 603  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 US-09-149-727-6

Query Match 51.0%; Score 1711; DB 3; Length 603;  
 Best Local Similarity 55.0%; Pred. No. 1.1e-133;  
 Matches 330; Conservative 93; Mismatches 161; Indels 16; Gaps 9;  
 40 VRPQRTSRELVLNDGLMKPAL--ASGLNDTAQPTAPLPKGLCEPVASYNDIFISRE 96  
 2 LRPEVPTREIKKLKLDGLMAFLDRENCIDQ--RWMESALOESRAIAVPSFNDQFADAD 59  
 97 IHDHVGWVYQREVEIVPKGWSQERLYVRASATHGRIVYNNRLVAEHVGYTPFEADVT 156  
 60 IRNYAGWVWYQREVEIVPKGWSQERLYVRASATHGRIVYNNRLVAEHVGYTPFEADVT 119  
 157 ELVAPGEKRLITGVNNELTHEITIPPKKITGTGATGKRIOTYQHDFFNYAGLARISWLYS 216  
 120 PYVIAKGSVRIITVCNNELMWQITIPGMVIT--DENGKKQSYPHDFPNYAGIHSVWLYT 178  
 217 VPQOHODITVYVDV--DGNGLINYEVEVANQTTGOIOISVIDEDGAIYAKASGAQGV 274  
 179 TPTWVDDITVTVHVAODCNHASVDWQV--VAN--GDVSEYLRDADQOVVATGQSTGTL 234  
 275 TIPSVKLMQFGAAYLYQLQVNIIVSSGDDVDTYNLATGVRTVYVAGSQPLINKPFFYTG 334  
 235 QVNNPHLMQGBEGYIVELCV--TAKSQTECDIPLRGIRSVAVKGBQPLINKPFFYTG 292  
 335 FGHEDTAVRGKHDPAVYVNDPQMKWIGANSFRTSHYVYAEVWDFADRNGLIVTDET 394  
 293 FGHEDADLRGKGFDDVNLVMDHADMWIGANSYRTSHYVYAEVWDFADRNGLIVTDET 352  
 395 PAYGLNIAL-MGVSSEBAP--QTFPTPAINDKTOEAKHQAIRELIARDKNASVVMMSIAN 452  
 353 AAVGFNLISLIGIEAGNKPEKLYSEEVNGETQAHQAIKELIARDKNPSVVMMSIAN 412  
 413 EPOTRPGAREYFAPRLAEATRKLDPTRPICVNVVFCDAHTDITISDLFDVLCINRYFGWY 472  
 513 SOTGDLEBAEALAEKELHGWQEKFRHPIVWTEYGADTLAGHSILGLPMSEEFQVOMLDM 572  
 473 VQSGDLETAKEVLEKELHGWQEKHOPITTEGVTTLAGHSIMYDMSEEFQVOMLDM 532



```

; PRIOR APPLICATION NUMBER: US 08/846,021
; PRIOR FILING DATE: 1997-04-25
; PRIOR APPLICATION NUMBER: US 08/366,783
; PRIOR FILING DATE: 1994-12-30
; PRIOR APPLICATION NUMBER: US 08/142,418
; PRIOR FILING DATE: 1993-11-16
; PRIOR APPLICATION NUMBER: US 07/659,835
; PRIOR FILING DATE: 1991-02-22
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37
; LENGTH: 604
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phae-GUS-phas
US-09-893-525-37

```

```

Query Match      50.9%; Score 1706; DB 4; Length 604;
Best Local Similarity 54.8%; Pred. No. 2.9e-133;
Matches 329; Conservative 94; Mismatches 161; Indels 16; Gaps 9;

```

```

QY 40 VRRPRTSRELVLNLDGLMKFAL--ASGLNDTAQPTAPLPGLECPVPASVNDIFISRE 96
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 3 LRPVETPTREIKKLDELMAFSLDRENGCIDQ--RMWESALQESRAIAVPSFNDQFADAD 60
QY 97 IHDHVMVYVYQREVIYPKMSQERLYVRAESATHHGRIVYNNRLVAEHVGGYTPFEADVT 156
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 IRNYAGNVVYQREVFIPKMGAGRIYLRFDVATHYGVWVNNQEVMEHQGYTPFEADVT 120
QY 157 ELVAPEKRLITGVNNELTHETIPPGKITTGATGKRIQTYOHDFFNYAGLARSITLYS 216
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 121 PYVIAKSVRIYVCNNELMWTIPPGMVIY--DENKKKQSYHDFPNYAGHRSVWLYT 179
QY 217 VPQOHIDITVTVDV--DGDNGLINYEVEVANOQTGOIISVIDEGALVAKASGAQTV 274
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 180 TPTNTWDDITVTVHVAODCNHASVDWQV--VAN--GDVSELRDADQOVVATGCGTSGLT 235
QY 275 TIPSVKLMQPGAAVLYLOQVNIYSSGDVVDYTNLATGVTVVAGSQFLINPKPFYFTG 334
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 236 QVNVPHLMQPGSELYVELCV--TAKSQTECDIYPLRAGRSVAVKQGFILNKPFFYFTG 293
QY 335 FGHGHEBTAVRGKGDPAVYVHDFOLMKWIGANSFRTSHYPAEEMDFADRNGLVIYDET 394
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 294 FGHHEBADIKRGKFDVNVHDMHADMWIGANSYRTSHYPAEEMDFADRNGLVIYDET 333
QY 395 PAVGNLIAL-MGVSESGAP-QFTTPDAINDKTOEAKHQAIRELIARDKXNASVVMMSIAN 452
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 354 AAVGFSLISIGIEAGNKPKELYSEAVNGETOAHLOAIKELIARDKXNASVVMMSIAN 413
QY 453 EPASHEDGAREYEPETNLTRQDPTPTFANVGATATYQDLRIISDLFDVSCINRYFGWY 512
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 414 EPDTRQAGAREYAPALAEATRKLDPTTRPICVMMFCDAHTDITISDLFDVLCINRYFGWY 473
QY 513 SQTGDLBEAFAALEKEKELHMQEKFHRPIYVTEYAGDTLAGLSILGLPMSSEFQOVQMLDM 572
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 474 VQSGDLETAKEVLEKELLMQEKLOHPIITTEYGVDTLAGLSHMYDMSSEFQOVQMLDM 533
QY 573 YHRVFDRIEMAGEHVMNFPADFOTNLGIIRVDGNKKGVFTDRKKPKRAAASLBARWTSID 632
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 534 YHRVFDRIEMAGEHVMNFPADFOTNLGIIRVDGNKKGVFTDRKKPKRAAASLBARWTSID 593

```

```

RESULT 8
; US-09-893-525-40
; Sequence 40, Application US/09893525
; Patent No. 6753167
; GENERAL INFORMATION:
; APPLICANT: Moloney, Maurice M.
; APPLICANT: Van Rooijen, GJjs
; TITLE OF INVENTION: Preparation of Heterologous Proteins on Oil Bodies
; FILE REFERENCE: 9369-172
; CURRENT APPLICATION NUMBER: US/09/893,525

```

```

; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 09/210,843
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: US 08/846,021
; PRIOR FILING DATE: 1997-04-25
; PRIOR APPLICATION NUMBER: US 08/366,783
; PRIOR FILING DATE: 1994-12-30
; PRIOR APPLICATION NUMBER: US 08/142,418
; PRIOR FILING DATE: 1993-11-16
; PRIOR APPLICATION NUMBER: US 07/659,835
; PRIOR FILING DATE: 1991-02-22
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40
; LENGTH: 659
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: phae-oleo GUS-phas
US-09-893-525-40

```

```

Query Match      50.9%; Score 1706; DB 4; Length 659;
Best Local Similarity 54.8%; Pred. No. 3.3e-133;
Matches 329; Conservative 94; Mismatches 161; Indels 16; Gaps 9;

```

```

QY 40 VRRPRTSRELVLNLDGLMKFAL--ASGLNDTAQPTAPLPGLECPVPASVNDIFISRE 96
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 58 LRPVETPTREIKKLDELMAFSLDRENGCIDQ--RMWESALQESRAIAVPSFNDQFADAD 115
QY 97 IHDHVMVYVYQREVIYPKMSQERLYVRAESATHHGRIVYNNRLVAEHVGGYTPFEADVT 156
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 116 IRNYAGNVVYQREVFIPKMGAGRIYLRFDVATHYGVWVNNQEVMEHQGYTPFEADVT 175
QY 157 ELVAPEKRLITGVNNELTHETIPPGKITTGATGKRIQTYOHDFFNYAGLARSITLYS 216
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 176 PYVIAKSVRIYVCNNELMWTIPPGMVIY--DENKKKQSYHDFPNYAGHRSVWLYT 234
QY 217 VPQOHIDITVTVDV--DGDNGLINYEVEVANOQTGOIISVIDEGALVAKASGAQTV 274
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 235 TPTNTWDDITVTVHVAODCNHASVDWQV--VAN--GDVSELRDADQOVVATGCGTSGLT 290
QY 275 TIPSVKLMQPGAAVLYLOQVNIYSSGDVVDYTNLATGVTVVAGSQFLINPKPFYFTG 334
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 291 QVNVPHLMQPGSELYVELCV--TAKSQTECDIYPLRAGRSVAVKQGFILNKPFFYFTG 348
QY 335 FGHGHEBTAVRGKGDPAVYVHDFOLMKWIGANSFRTSHYPAEEMDFADRNGLVIYDET 394
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 349 FGHHEBADIKRGKFDVNVHDMHADMWIGANSYRTSHYPAEEMDFADRNGLVIYDET 408
QY 395 PAVGNLIAL-MGVSESGAP-QFTTPDAINDKTOEAKHQAIRELIARDKXNASVVMMSIAN 452
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 409 AAVGFSLISIGIEAGNKPKELYSEAVNGETOAHLOAIKELIARDKXNASVVMMSIAN 468
QY 453 EPASHEDGAREYEPETNLTRQDPTPTFANVGATATYQDLRIISDLFDVSCINRYFGWY 512
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 469 EPDTRQAGAREYAPALAEATRKLDPTTRPICVMMFCDAHTDITISDLFDVLCINRYFGWY 528
QY 513 SQTGDLBEAFAALEKEKELHMQEKFHRPIYVTEYAGDTLAGLSILGLPMSSEFQOVQMLDM 572
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 529 VQSGDLETAKEVLEKELLMQEKLOHPIITTEYGVDTLAGLSHMYDMSSEFQOVQMLDM 588
QY 573 YHRVFDRIEMAGEHVMNFPADFOTNLGIIRVDGNKKGVFTDRKKPKRAAASLBARWTSID 632
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 589 YHRVFDRIEMAGEHVMNFPADFOTNLGIIRVDGNKKGVFTDRKKPKRAAASLBARWTSID 648

```

```

RESULT 9
; US-09-893-525-42
; Sequence 42, Application US/09893525
; Patent No. 6753167
; GENERAL INFORMATION:
; APPLICANT: Moloney, Maurice M.
; APPLICANT: Van Rooijen, GJjs

```



QY 513 SQTGDEAEAEAEKEHMOEKFRPIVWTEYAGDTLAGHSILGLPMSSEFQOVQMLDM 572  
 DB 472 VQSGDELEAKVLEKELAMQEKHQPIITTEYGVDTLAGHSMTYDMSSEFQOCAMLDM 531  
 QY 573 YHRVFDRISSMAGEHYWNPADPOTNIGIIRVDGNKKGVFTDRKPKAAHSLPARFTSID 632  
 DB 532 YHRVFDRISSMAGEHYWNPADPOTNIGIIRVDGNKKGVFTDRKPKAAHSLPARFTSID 591

## RESULT 11

US-09-151-957-5  
 ; Sequence 5, Application US/09151957  
 ; Patent No. 6429292

GENERAL INFORMATION:  
 APPLICANT: Jefferson, Richard A.  
 Wilson, Katherine J.

Leader: Michael

TITLE OF INVENTION: GLUCURONIDE REPRESSORS AND USES THEREOF

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SEED and BERRY LLP  
 STREET: 6300 Columbia Center, 701 Fifth Avenue  
 CITY: Seattle  
 STATE: Washington  
 COUNTRY: USA  
 ZIP: 98104-7092

COUNTRY: USA

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/151,957  
 FILING DATE: 11-Sep-1998  
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/882,704  
 FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:  
 NAME: No. 6429292tendburg Ph.D., Carol  
 REGISTRATION NUMBER: 39,317  
 REFERENCE/DOCKET NUMBER: 190106,404

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (206) 622-4900  
 TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 602 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
 US-09-151-957-5

Query Match 49.6%; Score 1664.5; DB 4; Length 602;  
 Best Local Similarity 54.0%; Pred. No. 8,2e-130;  
 Matches 324; Conservative 93; Mismatches 166; Indels 17; Gaps 10;

QY 40 VRPQRTSSRELVLNLDGLMKFAL--ASGLNDTAQPTAPLPKGLCECPVSPASYNDIFISRE 96  
 DB 2 LRPEVETPREIKKLDGLMAFSLDRENGCIDQ--RWMESALQESRAIAVGSFNDQFADAD 59  
 QY 97 IHDHVMVYVQREVIYVPGKMSQERIVLRASATTHGRIVYNNRLVAEHVGGYTPFEADVT 156  
 DB 60 IRNYAGVWVYQREVIYVPGKMSQERIVLRASATTHGRIVYNNRLVAEHVGGYTPFEADVT 119  
 QY 157 ELVAPEKEFLITGVNNELTHEITIPPKITITGNATGKRIQTOYHDFYVYAGLARSITWLS 216  
 DB 120 PYVIAGKSVRITVCVNNELMWOTIIPPGWIT--DENKKKQSYTHDFPNYAGIRSWLTYT 178  
 QY 217 VPQOHIQDITVTVTV--DGDNGLINYEVEVANOQTGOIQISVIDEGAIYAKASGAQGTV 274  
 DB 179 TPRTWVDITVTVTHVADQCNHASVDMQV--VAN---GDVSEVELDADQOVVATGQTSGLT 234

QY 275 TIPSVKLMQPGAAVLYLOLVNIVGSSGDVDTYNLATGVRTVYVAGSQFLINCKPEPYFTG 334  
 DB 235 QVNVPHLMQPGESYLYELCV--TAKSQTECDIYPLVAGRSVAVKGEQFLINCKPEPYFTG 292  
 QY 335 FGRHEDTAVRGKGDPAVYVHDFOLMKWIGANSFRTSHYPYAEVMDPADRNGIIVYDET 394  
 DB 293 FGRHEDTAVRGKGDPAVYVHDFOLMKWIGANSFRTSHYPYAEVMDPADRNGIIVYDET 352  
 QY 395 PAVGNIAL--MGVSESGAP--QFTTPAINDKTOEAKHAKIRELIADKXNASVYVMSIAN 452  
 DB 353 AAVGNLSLIGIEAGNKKRELSEEAUVNGEQAHLOAIKELIADKXNASVYVMSIAN 412  
 QY 453 EPASHEDGAREYEPETNLTRQDPPRTPTFANVATATVQDLRISDLPVSCINRYFGWY 512  
 DB 413 EPPTRPQ--VHGNISPLAEATKIDPRTPTTCVNVVCDANTDITSDLPVCLNRYFGWY 471

QY 513 SQTGDEAEAEAEKEHMOEKFRPIVWTEYAGDTLAGHSILGLPMSSEFQOVQMLDM 572  
 DB 472 VQSGDELEAKVLEKELAMQEKHQPIITTEYGVDTLAGHSMTYDMSSEFQOCAMLDM 531  
 QY 573 YHRVFDRISSMAGEHYWNPADPOTNIGIIRVDGNKKGVFTDRKPKAAHSLPARFTSID 632  
 DB 532 YHRVFDRISSMAGEHYWNPADPOTNIGIIRVDGNKKGVFTDRKPKAAHSLPARFTSID 591

## RESULT 12

5432081-2  
 ; Patent No. 5432081

APPLICANT: JEFFERSON, RICHARD A.  
 TITLE OF INVENTION: HOST CELLS TRANSFORMED WITH THE E. COLI  
 ; GLUCURONIDE PERMEASE GENE

NUMBER OF SEQUENCES: 10  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/138,546  
 FILING DATE: 15-OCT-1993

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 447,976  
 FILING DATE: 08-DEC-1989

APPLICATION NUMBER: 264,586  
 FILING DATE: 31-OCT-1988

APPLICATION NUMBER: 119,102  
 FILING DATE: 10-NOV-1987

SEQ ID NO: 2:  
 LENGTH: 602

Query Match 49.6%; Score 1664.5; DB 6; Length 602;  
 Best Local Similarity 54.0%; Pred. No. 8,2e-130;  
 Matches 324; Conservative 93; Mismatches 166; Indels 17; Gaps 10;

QY 40 VRPQRTSSRELVLNLDGLMKFAL--ASGLNDTAQPTAPLPKGLCECPVSPASYNDIFISRE 96  
 DB 2 LRPEVETPREIKKLDGLMAFSLDRENGCIDQ--RWMESALQESRAIAVGSFNDQFADAD 59  
 QY 97 IHDHVMVYVQREVIYVPGKMSQERIVLRASATTHGRIVYNNRLVAEHVGGYTPFEADVT 156  
 DB 60 IRNYAGVWVYQREVIYVPGKMSQERIVLRASATTHGRIVYNNRLVAEHVGGYTPFEADVT 119  
 QY 157 ELVAPEKEFLITGVNNELTHEITIPPKITITGNATGKRIQTOYHDFYVYAGLARSITWLS 216  
 DB 120 PYVIAGKSVRITVCVNNELMWOTIIPPGWIT--DENKKKQSYTHDFPNYAGIRSWLTYT 178  
 QY 217 VPQOHIQDITVTVTV--DGDNGLINYEVEVANOQTGOIQISVIDEGAIYAKASGAQGTV 274  
 DB 179 TPRTWVDITVTVTHVADQCNHASVDMQV--VAN---GDVSEVELDADQOVVATGQTSGLT 234  
 QY 275 TIPSVKLMQPGAAVLYLOLVNIVGSSGDVDTYNLATGVRTVYVAGSQFLINCKPEPYFTG 334  
 DB 235 QVNVPHLMQPGESYLYELCV--TAKSQTECDIYPLVAGRSVAVKGEQFLINCKPEPYFTG 292  
 QY 335 FGRHEDTAVRGKGDPAVYVHDFOLMKWIGANSFRTSHYPYAEVMDPADRNGIIVYDET 394  
 DB 293 FGRHEDTAVRGKGDPAVYVHDFOLMKWIGANSFRTSHYPYAEVMDPADRNGIIVYDET 352

Db 293 FGRHEDADLRKGFQDNLVMDHMDLMDIGANSYRSHYPAEMLDWADEHGIIVIDET 352  
 QY 395 PAVGNIAL-MGVSESGAP-QTFPPDAINDKTOEAKHQAIRELIARDKNHASYVMWSIAN 452  
 Db 353 AAVGFNLISLIGIFGPAKPKRELYSEEAANGETOQAHLQAIKELIARDKNHASYVMWSIAN 412  
 QY 453 EPASHEDGARREYEPFLTNLRQDLPTRPITFANVGATATYQDLRISDLFVSCINRYFGWY 512  
 Db 413 EPDTRPQ-VHGNISPLAEATRKLDPTRPITCVNMFCDANTDTISDLFVCLNRYGYM 471  
 QY 513 SQTGDLLEAEALAEKELHGMQEKFRPIVWTEYGADTLAHLISILGLPMSEEFQVOMLDM 572  
 Db 472 VQSGDLETAKEVLEKELLMQEKHOPIIITEYGVDTLGLHSMWTDMSESEYQCAWLDM 531  
 QY 573 YHRVFDRIESMAGEHWNFNADPOTNIGIIRVDGNKKGVFTDRDKPKAAHSLRARWTSID 632  
 Db 532 YHRVFDRIESMAGEHWNFNADPOTNIGIIRVDGNKKGVFTDRDKPKAAHSLRARWTSID 591

## RESULT 13

5432081-2

Patent No. 5432081  
 APPLICANT: JEFFERSON, RICHARD A.  
 TITLE OF INVENTION: HOST CELLS TRANSFORMED WITH THE E. COLI  
 GLUCONONIDE PERMEASE GENE

NUMBER OF SEQUENCES: 10  
 CURRENT APPLICATION NUMBER: 10

FILING DATE: 15-OCT-1993  
 PRIOR APPLICATION NUMBER: 447,976

FILING DATE: 08-DEC-1989  
 APPLICATION NUMBER: 264,586

FILING DATE: 31-OCT-1988  
 APPLICATION NUMBER: 119,102

FILING DATE: 10-NOV-1987  
 SEQ ID NO: 2

LENGTH: 602

5432081-2

Query Match 49.6%; Score 1664.5; DB 6; Length 602;  
 Best Local Similarity 54.0%; Pred. No. 8.2e-130;

Matches 344; Conservative 93; Mismatches 166; Indels 17; Gaps 10;

QY 40 VREPORTSRELVLNDGLKFPAL--ASGLNDTAQWTPAPLPKGLCEPVASVNDIFISRE 96  
 Db 2 LRVEPTREIKKLDGLMAFSLDRENCIDQ--RWMSALQESRAIAPVGSFNQFADAD 59  
 QY 97 IDHNVGWVYVOREVIVPKGMSQERYLVRAESATHGRIYVNNRLVAEHVGITPFEADV 156  
 Db 60 IRNYAGNVVYOREVIVPKGMSQERYLVRAESATHGRIYVNNRLVAEHVGITPFEADV 119  
 QY 157 ELVAPGEKRLITGVNNELTHEITPPKITTGNATGKRIQTYQHDFYVYAGLARSIMLYS 216  
 Db 120 PYVIAGKSVRITVCNNELMQITIPPGWIT--DENGKKQSYFHFENYAGIHSVIMLYT 178  
 QY 217 VPQOHIQDITVVDV--DGNGLINYEVEVANTQGOIISVIDEGAIIVAKASAGQTV 274  
 Db 179 TPNTWDDITVTVHVDQCNHASVDMQV-VAN--GDVSELRLDADQVAVATGQTSGLT 234  
 QY 275 TIPSVKLMOPGAALYLOLVNIVSSGDVVDYTNLATGVRTYKVASGQFLNGKPFYFTG 334  
 Db 235 QVAVPHLMOPGEGIVLELV--TAKSQTECDIYPLRAGIRSVAVKGEOLINHKPFYFTG 292  
 QY 335 FGRHEDTAVRGKGDPAVMYNDPQLMKWTGANSFRTSHYPAEEMLDWADDEHGIIVIDET 394  
 Db 293 FGRHEDADLRKGFQDNLVMDHMDLMDIGANSYRSHYPAEEMLDWADDEHGIIVIDET 352  
 QY 395 PAVGNIAL-MGVSESGAP-QTFPPDAINDKTOEAKHQAIRELIARDKNHASYVMWSIAN 452  
 Db 353 AAVGFNLISLIGIFGPAKPKRELYSEEAANGETOQAHLQAIKELIARDKNHASYVMWSIAN 412  
 QY 453 EPASHEDGARREYEPFLTNLRQDLPTRPITFANVGATATYQDLRISDLFVSCINRYFGWY 512

Db 413 EPDTRPQ-VHGNISPLAEATRKLDPTRPITCVNMFCDANTDTISDLFVCLNRYGYM 471  
 QY 513 SQTGDLLEAEALAEKELHGMQEKFRPIVWTEYGADTLAHLISILGLPMSEEFQVOMLDM 572  
 Db 472 VQSGDLETAKEVLEKELLMQEKHOPIIITEYGVDTLGLHSMWTDMSESEYQCAWLDM 531  
 QY 573 YHRVFDRIESMAGEHWNFNADPOTNIGIIRVDGNKKGVFTDRDKPKAAHSLRARWTSID 632  
 Db 532 YHRVFDRIESMAGEHWNFNADPOTNIGIIRVDGNKKGVFTDRDKPKAAHSLRARWTSID 591

## RESULT 14

US-09-488-270A-2

Sequence 2, Application US/09488270A  
 Patent No. 6455759

GENERAL INFORMATION:  
 APPLICANT: Vilestra, Richard D

APPLICANT: Walker, Joseph M  
 TITLE OF INVENTION: Production of Multiple Proteins in Plants

FILE REFERENCE: 960296.96501  
 CURRENT FILING DATE: 2000-01-20

NUMBER OF SEQ ID NOS: 4  
 SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 2

LENGTH: 1242

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE: Description of Artificial Sequence: Fusion Protein

OTHER INFORMATION: Expression Cassette

US-09-488-270A-2

Query Match 49.6%; Score 1664.5; DB 4; Length 1242;  
 Best Local Similarity 54.0%; Pred. No. 2.5e-129;

Matches 324; Conservative 93; Mismatches 166; Indels 17; Gaps 10;

QY 40 VREPORTSRELVLNDGLKFPAL--ASGLNDTAQWTPAPLPKGLCEPVASVNDIFISRE 96  
 Db 642 LRVEPTREIKKLDGLMAFSLDRENCIDQ--RWMSALQESRAIAPVGSFNQFADAD 699  
 QY 97 IDHNVGWVYVOREVIVPKGMSQERYLVRAESATHGRIYVNNRLVAEHVGITPFEADV 156  
 Db 700 IRNYAGNVVYOREVIVPKGMSQERYLVRAESATHGRIYVNNRLVAEHVGITPFEADV 759  
 QY 157 ELVAPGEKRLITGVNNELTHEITPPKITTGNATGKRIQTYQHDFYVYAGLARSIMLYS 216  
 Db 760 PYVIAGKSVRITVCNNELMQITIPPGWIT--DENGKKQSYFHFENYAGIHSVIMLYT 818  
 QY 217 VPQOHIQDITVVDV--DGNGLINYEVEVANTQGOIISVIDEGAIIVAKASAGQTV 274  
 Db 819 TPNTWDDITVTVHVDQCNHASVDMQV-VAN--GDVSELRLDADQVAVATGQTSGLT 874  
 QY 275 TIPSVKLMOPGAALYLOLVNIVSSGDVVDYTNLATGVRTYKVASGQFLNGKPFYFTG 334  
 Db 875 QVAVPHLMOPGEGIVLELV--TAKSQTECDIYPLRAGIRSVAVKGEOLINHKPFYFTG 932  
 QY 335 FGRHEDTAVRGKGDPAVMYNDPQLMKWTGANSFRTSHYPAEEMLDWADDEHGIIVIDET 394  
 Db 933 FGRHEDADLRKGFQDNLVMDHMDLMDIGANSYRSHYPAEEMLDWADDEHGIIVIDET 352  
 QY 395 PAVGNIAL-MGVSESGAP-QTFPPDAINDKTOEAKHQAIRELIARDKNHASYVMWSIAN 452  
 Db 993 AAVGFNLISLIGIFGPAKPKRELYSEEAANGETOQAHLQAIKELIARDKNHASYVMWSIAN 412  
 QY 453 EPASHEDGARREYEPFLTNLRQDLPTRPITFANVGATATYQDLRISDLFVSCINRYFGWY 512  
 Db 1053 EPDTRPQ-VHGNISPLAEATRKLDPTRPITCVNMFCDANTDTISDLFVCLNRYGYM 1111  
 QY 513 SQTGDLLEAEALAEKELHGMQEKFRPIVWTEYGADTLAHLISILGLPMSEEFQVOMLDM 572  
 Db 532 YHRVFDRIESMAGEHWNFNADPOTNIGIIRVDGNKKGVFTDRDKPKAAHSLRARWTSID 591  
 QY 1112 VQSGDLETAKEVLEKELLMQEKHOPIIITEYGVDTLGLHSMWTDMSESEYQCAWLDM 1171



Qy 573 YHRVFDRIESMAGEHYNFPADFOTNLGIRVDGNKKGVFTDRKPKAAASHLRARWTSID 632  
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :  
Db 1172 YHRVFDRIESMAGEHYNFPADFOTNLGIRVDGNKKGVFTDRKPKAAASHLRARWTSID 1231

RESULT 15  
5268463-2  
; Patent No. 5268463  
; APPLICANT: JEFFERSON, RICHARD A.  
; TITLE OF INVENTION: PLANT PROMOTER a-GLUCURONIDASE GENE  
; CONSTRUCT  
; NUMBER OF SEQUENCES: 9  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/447,976  
; FILING DATE: 08-DEC-1989  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: 119,102  
; FILING DATE: 10-NOV-1987  
; APPLICATION NUMBER: 264,586  
; FILING DATE: 31-OCT-1988  
; SEQ ID NO: 2:  
; LENGTH: 600  
5268463-2

Query Match 48.9%; Score 1638.5; DB 6; Length 600;

Best Local Similarity 53.8%; Pred. No. 1.2e-127; Matches 333; Conservative 93; Mismatches 165; Indels 19; Gaps 12;

Qy 40 VREPORTSRELVNLDGIMKFAAL--ASGLNDTAQPTAPLPKGLCECPVPAASYNDIFISRE 96  
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :  
Db 2 LRVERPTREIKLDGLMAFSLDRENCIDQ--RWMSALQESRAIAPESFNDQFADAD 59  
Qy 97 IHDVGVWVYQREVIYPKGWSQERYLVKASATHHGRIVYNNRLVAHVGGYTFEADVT 156  
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :  
Db 60 IRNYAGNVWYQREVIYPKGWSQERYLVKASATHHGRIVYNNRLVAHVGGYTFEADVT 119  
Qy 157 ELVAPGEKRLITGVNNELTHETIPPKKITGTGATGRIQTYQHDPFYNAVGLARSIMLYS 216  
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :  
Db 120 PYVIAGKSVRIITCVNNELMWQITTPGMVTT--DENGKKOSYFHDFFNAGIHRSMVLT 178  
Qy 217 VPOOHIDITVTVDV--DGDNGLINYEVEVANQTTGQIQISVIDEDGAIYAKASGAQTV 274  
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :  
Db 179 TPRTWVDITTVVTHVADCNHASVDNQV--VAN--GDVSELRDADQVAVATGQTSGLT 234  
Qy 275 TIPSVLKQPGAAVYVQLOVNISSGDDVDVYTNLATGVRTVYAGSQFLINGKPPYFTG 334  
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :  
Db 235 QVNPPLMOPG--EYLVELCV--TAKSQTECDIPLRVGIRSVAVKGEQFLINHKPFYFTG 291  
Qy 335 FGHEDTAVRGKHDPAVWVHDQMLKMTGANSFRTSHYPAEEVDFADRNQIVVIDET 394  
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :  
Db 292 FGHEDADIRGKFDVNLVHDHALMDWIGANSYRTSHYPAEEVDFADRNQIVVIDET 351  
Qy 395 PAVGLNIAL--MGVSESGAP--QFTPDALNDKTOEAKHQAIRELIARDKNHASVWMSIAN 452  
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :  
Db 352 AAVGFNLSIGIRFEAGNKREKELYSEBAVNGETOAHLOAKELIARDKNHPSVWMSIAN 411  
Qy 453 EPASHEDGAREYFEPLTNLTROLDPTRPITFANVGTATYQDRIIDLFDVSCINRYFGWY 512  
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :  
Db 412 EPDTRPQ--VHGNISPLAEATRKLDPTRPITCVNMFCDATDTISDLFDVLCINRYFGWY 470  
Qy 513 SQTGDLFEAALAEKELHGOEKFRPIVWTEGADTLAGHSILGLPWSEEFQVOMLDM 572  
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :  
Db 471 VOSGDLETAKVLEKELAMQERKHQPIITTEYGVDTLAGLHSMYTDWSESEYQCAMLDM 530  
Qy 573 YHRVFDRIESMAGEHYNFPADFOTNLGIRVDGNKKGVFTDRKPKAAASHLRARWTSID 632  
| | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :  
Db 531 YHRVFDRIESMAGEHYNFPADFOTNLGIRVDGNKKGVFTDRKPKAAASHLRARWTSID 589

Search completed: March 18, 2005, 23:53:55  
Job time : 46 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

# OM protein - protein search, using SW model

Run on: March 18, 2005, 23:33:12 ; Search time 171 Seconds  
(without alignments)  
1433.954 Million cell updates/sec

Title: US-10-757-093-4  
Perfect score: 3354  
Sequence: 1 MKFLTGLSLSLAAPSIGTP.....RKPKAAHSLRAWTSIDKN 634

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues  
Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1756	52.4	602	3	AAB28409
2	1714.5	51.1	832	2	AAW04302
3	1714	51.1	618	7	ADL01666
4	1714	51.1	1010	3	AAV68840
5	1711	51.0	603	2	AAW93827
6	1711	51.0	603	2	AAW93824
7	1711	51.0	603	2	AAW93824
8	1711	51.0	603	5	ABW84108
9	1711	51.0	603	6	ABW96657
10	1711	51.0	603	8	ABU19649
11	1711	51.0	603	8	ADN18189
12	1706	50.9	604	7	ADD27986
13	1706	50.9	659	7	ADD27989
14	1706	50.9	850	7	ADD27991
15	1697	50.6	603	5	ABW84107
16	1667.5	49.7	602	2	AAW43387
17	1665.5	49.6	607	8	ADS26258
18	1665	49.6	711	6	ABR83626
19	1664.5	49.6	602	2	AAW42429
20	1664.5	49.6	1242	5	ABW81108
21	1664.5	49.6	1242	6	ABW84637
22	1657.5	49.4	602	1	AAW82948
23	1590.5	47.4	617	4	AAU39683
24	1590.5	47.4	617	6	ABW36202
25	1439	42.9	615	3	AAB28408

26	1439	42.9	618	2	AAW93821	AAW93821 Bacillus
27	1437	42.8	602	2	AAW93825	AAW93825 Bacillus
28	1437	42.8	602	2	AAW93822	AAW93822 Bacillus
29	1437	42.8	602	2	AAW93826	AAW93826 Bacillus
30	1437	42.7	602	3	AAW28402	AAW28402 Staphyloc
31	1431	42.4	602	3	AAW93820	AAW93820 Bacillus
32	1338	39.9	613	2	AAW93823	AAW93823 E. coli G
33	1338	39.9	613	2	AAW93828	AAW93828 Human GUS
34	1338	39.9	613	3	AAW28407	AAW28407 Escherich
35	1338	39.9	613	4	AAW62276	AAW62276 Mutant he
36	1338	39.9	613	4	AAW62271	AAW62271 Heavy cha
37	1338	39.9	651	4	AAW02443	AAW02443 Human bet
38	1338	39.9	651	7	ADW45481	ADW45481 Human pro
39	1338	39.9	651	7	ADW57448	ADW57448 Human pro
40	1338	39.9	651	8	ADP12392	ADP12392 Protein e
41	1338	39.9	651	8	ADW98808	ADW98808 Antagonis
42	1338	39.9	722	6	AAW33322	AAW33322 L. mexica
43	1338	39.9	722	8	ADW47503	ADW47503 Human bet
44	1338	39.9	722	8	ADW58609	ADW58609 Fusion pr
45	1335	39.8	648	7	ADW45479	ADW45479 Rat Prote

## ALIGNMENTS

RESULT 1  
AAB28409  
ID AAB28409 standard; protein; 602 AA.  
XX  
AC AAB28409;  
XX  
DT 26-JAN-2001 (first entry)  
XX  
DE Salmoneilla beta-glucuronidase.  
XX  
KM Microbial; beta-glucuronidase; GUS; Enterobacter; Salmoneilla;  
KW Pseudomonas; Staphylococcus; Thermotoga; transgenic plant; bioindicator;  
KW transgenic insect; marker; glucuronide detoxification.  
XX  
OS Salmoneilla sp.  
XX  
PN WO200055333-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 16-MAR-2000; 2000WO-US007107.  
XX  
PR 17-MAR-1999; 99US-00270957.  
XX  
PA (CAMP-) CAMBIA BIOSYSTEMS LLC.  
XX  
PI Jefferson RA, Mayer JE;  
XX  
DR WPI: 2000-647075/62.  
XX  
DR N-PSDB; AAA07939.  
XX  
PT Novel microbial beta-glucuronidase genes and gene products used as  
PT reporter/effector molecule, as diagnostic tool, in positive selection, to  
PT target molecules to specific cells and to detect and track linked genes.  
XX  
PS Example 3; Fig 17; 116pp; English.  
XX  
CC The present sequence is a microbial beta-glucuronidase (GUS) protein. GUS  
CC genes were obtained from six different genera: Enterobacter/Salmoneilla,  
CC Pseudomonas, Salmoneilla, Staphylococcus and Thermotoga. Microbial GUS can  
CC be used as a reporter/effector molecule for transgenic constructions and  
CC in vitro diagnostic applications. It may also be used to generate  
CC sentinel plants that serve as bioindicators of environmental status. It  
CC may be used to generate transgenic insects for tracking insect  
CC populations or to facilitate the development of a bioassay for compounds  
CC that affect molecules critical for insect development (e.g. juvenile  
CC hormone). Secreted GUS may also serve as a marker for beneficial fungi  
CC destined for release into the environment. In animal systems, secreted

CC GUS may be used to achieve extracellular detoxification of glucuronides  
 CC (e.g. toxin glucuronide) and to examine conjugation patterns of  
 CC glucuronides. Microbial GUS may also be used in traditional medical  
 CC diagnostic assays, for drug testing, pharmacokinetic studies,  
 CC bioavailability studies, diagnosis of diseases and syndromes, following  
 CC progression of disease or its response to therapy. Microbial GUS has  
 CC increased thermal stability, high turnover number and enzymatic activity.  
 CC It is highly specific for the substrate and water soluble, and the  
 CC substrates are stable

XX Sequence 602 AA;

Query Match 52.4%; Score 1756; DB 3; Length 602;

Best Local Similarity 55.1%; Pred. No. 3.8e-134;  
 Matches 329; Conservative 99; Mismatches 159; Indels 10; Gaps 6;

QY 40 VRPQRSSRLVNDGKMFALASGLNDTQP-WTAPLPGLGCPYPASINDIFISETH 98  
 DB 2 LRSVETATREIKKLDGMSFCMDSBECGNAQMMWRQPLPSRAIIVGSGSYNDQFAAEIR 61  
 QY 99 DHVGVVYQREVIIVPKGMSQERYLVRAESATHHGRIVNNRLVAEHVGYTPPEADVTTEL 158  
 DB 62 NYGVNWWYQREIRIPKMDRQRIYLRFDATYHGKWNVDQFLMEHQGTFPEADISHL 121  
 QY 159 VADGEKRLTGVNNELTHETTPPKITGNATGKRIQYQHPYVAGIARSIMLYSV 218  
 DB 122 ISAGESVRIIVCVNNELMTQITPPGVVYQG-VNGKKQAFYHFFVYAGIHSRIMLYTTP 180  
 QY 219 QGHIDITVVDGNGNLNVEVANQTTQIQISVIDEDGAIYAKSAGACTYTPS 278  
 DB 181 KTRVEDITVTVQVADD-LAQATVAMOVNANGVRRLAEQOLVASGGEGKGLLEG 238  
 QY 279 VKLMQGAAYLYQLVNIVGSSGDVDTYNLATGVRTVKVASQFLINGKPYFTGFGKH 338  
 DB 239 PRMVGEGVLYELRV-IAQHQBDEDEYELRGISVEVKGQFLINHPFTFTGGRH 296  
 QY 339 EDPAVNGKGDPAVWHDFOIMKIGANSFRITSHYPAEEMDPADRNGLIVIDEFPANG 398  
 DB 297 EDADLKGKGDVNLVMDHMLDMIGANSYRISHYPAEEMDLMADEHGIVIIDETPAVNG 356  
 QY 399 LNTALNGVS--BSGAPQPTFPAINDKTOEAKHQAIRELIARDKNHASVVMSSIANEPA 455  
 DB 357 FNISSL-GISDVGEKREKELYSDEAVNDETQRAHLOAKELIARDKNHPSVVMSSIANEPD 415  
 QY 456 SHEDGAREYFEPULNTRQLDPTRPITTFANVGTATYQDLRISDLPVSCINRYFGWSQT 515  
 DB 416 TRNNGAREYFAPLAQTRELDPTRPITTCVVMFCDAESDTITDLPVVCINRYFGWSQT 475  
 QY 516 GDLEEAALKEKLEHGMQEKFNPIYWTYRGADTLAAGLSILGLPWSEEPQVOMLDMYHR 575  
 DB 476 GDLEKAKEKLEKELAMOEKLRPIITTEGYVDTLAAGLSHMYNDMWSSEEQCAMLDMYHR 535  
 QY 576 VPDRISSMAGEHWNPADPOTNLGITRVQDNKGVTPDRKPKAAHSLPARVTSID 632  
 DB 536 VPDRISSAVNGEYQWNPADPOTNLGITRVQDNKGVTPDRKPKAAHSLPARVTSID 592  
 RESULT 2  
 ID AAM04302  
 XX AAM04302 standard; protein; 832 AA.  
 AC AAM04302;  
 XX  
 XX 25-MAR-2003 (revised)  
 DT 16-FEB-1997 (first entry)  
 XX  
 DE Antibody/beta glucuronidase fusion protein.  
 XX  
 XX Antibody; fusion protein; recombinant antibody; tumour therapy; prodrug.  
 OS Synthetic.  
 XX  
 PN EPJ73747-AA.

XX  
 PD 16-OCT-1996.  
 XX  
 PF 13-MAR-1996; 96EP-00103913.  
 XX  
 PR 11-APR-1995; 95DE-01013676.  
 XX  
 PA (BEHN) BEHRINGMERKE AG.  
 XX  
 PI Oper M, Bosajet K, Czech J;  
 XX  
 XX WPI; 1996-457328/46.  
 DR N-PSDB; AAT18397.  
 XX  
 PT Prod. of recombinant antibody (Ab), Ab fragment or Ab/enzyme fusion  
 PT protein - by cytoplasmic expression in thio:reductoxin:reductase deficient  
 PT E. coli.  
 PS  
 XX Example 1; Fig 5a; 12pp; German.

CC Production of recombinant antibodies (Ab), Ab fragments or Ab  
 CC fragment/enzyme fusion proteins can be used for tumour therapy,  
 CC especially when the fusion protein comprises a tumour-specific Ab  
 CC fragment and an enzyme capable of converting a non-toxic prodrug to a  
 CC toxic drug. The fusion proteins are constructed in expression vectors and  
 CC expressed in thio:reductoxin:reductase deficient E.coli, allowing expression  
 CC products to be isolated in soluble functional form without renaturation.  
 CC The Ab fragment is an Fab fragment or an antigen binding region. In the  
 CC fusion protein, the Ab component is humanised and the enzyme component is  
 CC a human cytoplasmic enzyme. This fusion protein comprises the antibody  
 CC constant and variable heavy chain regions and the E.coli Beta  
 CC glucuronidase enzyme. (Updated on 25-MAR-2003 to correct PR field.)  
 XX

SO Sequence 832 AA;

Query Match 51.1%; Score 1714.5; DB 2; Length 832;

Best Local Similarity 52.7%; Pred. No. 1.5e-130;  
 Matches 338; Conservative 98; Mismatches 180; Indels 25; Gaps 11;

QY 8 SLLSLAAPSGLTPA-----ARHPRNEMTQHEQPL-----IKVRPQRTSRLVNDGLMK 58  
 DB 190 SVVTVPPSSSLGTQYTCNNVHKRPSTKYDKRYELSGSSGSMVRPVTTPPEIKKLDGLMA 249  
 QY 59 FAL--ASGLNDTAPQWAPLPKGLGECVPYASINDIFISREIHADVGVVYQREVIIVPKG 115  
 DB 250 FSLDRENGCIDQ--RWMSALQESRAIIVPGSFNDQFADADIRVAGVNWYQREVIIVPKG 307  
 QY 116 NSQERYLVASASATHHGRIVYNNRLVAEHVGYTTPPEADVTLELVADGEKRLTGVNNEL 175  
 DB 308 MAGQRIVLRFDAVTHYGVKVVANNOEVMERHOGYTTPPEADVTLELVAGKSVRTVCNNEL 367  
 QY 176 THEITPPGKITGNATGKRIQYQHPYVAGIARSIMLYSVPOQHIDITVTVQV--DG 233  
 DB 368 NMOTITPPGVNIT--DENGKKKQSYFNFYPAISHSVMLYTPENWVDDITVTVHAQDC 426  
 QY 234 DNGLINYEVEVANQTTQIQISVIDEDGAIYAKSAGACTYVTPSVKLMQGAAYLYQLO 293  
 DB 427 NHASVDQWV-VAN--GDVSVELRDADQGVVATGQGTGGLQVNPRLMQPSEGVLYELC 482  
 QY 294 VNIIVSSGDVDTYNLATGVRTVKVASQFLINGKRPFTFTGKGKEDPAVNRKGDPAVM 353  
 DB 483 V--TAKSQTECDIYPLRGVIRSVAAVKGQFLINHPFPFTGGRHEDADLRKKGKPDNYLM 540  
 QY 354 VHDFOIMKIGANSFRITSHYPAEEMDPADRNGLIVIDEFPANGVNLAL-MGVSESGAP 412  
 DB 541 VHDHMLDMIGANSFRITSHYPAEEMDLMADEHGIVVIDETPAVNFSLGIGFAGAKNP 600  
 QY 413 -QTFPDAINDTQAHQAIRELIARDKNHASVVMSSIANEPASHEGAREYFEPULNL 471  
 DB 601 KELYSEAVNGEYQWNPADPOTNLGITRVQDNKGVTPDRKPKAAHSLPARVTSID 660  
 QY 472 TRQLDPTRPITTFANVGTATYQDLRISDLPVSCINRYFGWSQTDLEEAALKEKLEHG 531

DB 661 TRKLDTPTPTCVNMFCDNHTDITSDLPVLCINRYGYVSGDLETAKEVLEKELLA 720  
 QY 532 MOSKPRPIVMEYAGDTLAGLSILGLPKSEEPQVOMLMYHRVPRIRISMGHEHWNF 591  
 DB 721 MOSKLPPIITITTYGDTLAGLSMTYDMWSEEQCAMLDMYHRVPRVSAVVGEOVWNF 780  
 QY 592 ADEFQNIIGIIRVDGNKKGVFTDRDKPKYAAHSLRARWTSID 632  
 DB 781 ADFATSGILRVGNGKKGIPTDRDKPKSAFLQKRTGKN 821

RESULT 3  
 ADL01666  
 ID ADL01666 standard; protein; 618 AA.  
 XX

ADL01666:  
 06-MAY-2004 (first entry)

Modified GUS protein HGUH amino acid sequence.

XX GUS; beta-glucuronidase; fusion polypeptide; Extn; Extc; Int;  
 XX N-terminal portion; Intein; C-terminal portion; plant optimised codon;  
 KM protein splicing mechanism; transgene;  
 KM multi-functional hybrid protein polymer; circular protein.

XX Unidentified.  
 OS Synthetic.

PN WO2003066861-A1.

PD 14-AUG-2003.

PF 04-FEB-2003; 2003WO-US003435.

PR 04-FEB-2002; 2002US-0354395P.

PA (DUPO ) DU PONT DE NEMOURS & CO E I.

PI Yadav NS, Yang JG;

DR WPI; 2003-731504/69.

PT New isolated polynucleotide encoding a polypeptide comprising an Extn, an  
 PT Extc, and an Int interspersed between the Extn and the Extc, useful for  
 PT Intein-mediated protein splicing.

PS Example 2; SEQ ID NO 28; 63bp; English.

XX This invention relates to a novel isolated polynucleotide comprising a  
 CC nucleotide sequence that encodes a polypeptide or a fusion polypeptide  
 CC comprising an Extn, an Extc and an Int interspersed between the Extn and  
 CC the Extc. The Extn is the N-terminal portion of the polypeptide, the Int  
 CC is an intein, and the Extc is the C-terminal portion of the polypeptide.  
 CC At least a portion of the nucleotide sequence has been modified to  
 CC contain plant optimised codons. The polynucleotide and methods are useful  
 CC in introducing a protein splicing mechanism into plants by employing  
 CC inteins and transgenes. This permits in vivo and in vitro synthesis of  
 CC homogeneous and large multi-functional hybrid protein polymers and  
 CC circular proteins. The present sequence is that of a protein which is  
 CC related to the invention.

XX Sequence 618 AA;

Query Match 51.1%; Score 1714; DB 7; Length 618;  
 Best Local Similarity 55.2%; Pred. No. 1.1e-130;  
 Matches 331; Conservative 92; Mismatches 161; Indels 16; Gaps 9;

QY 40 VRQQRSSRLVNDGIMKFAAL--ASGLNDTAQPTAPPKGLGCVYPSYNDIFISRE 96  
 DB 9 VRPEVPTTRBIKLDGIMAFSLDRENGIDQ--RWMSALQESRAIVPDSFNDQFAD 66  
 QY 97 IHGVGVVYQREVIYVKGMSQERYLVRAVSATHHGRIVYNNRLVAEHVGQYTFEADVT 156

DB 67 IRNAYGVWVYQREVFIPKGNAGQRIYLRFDAVTHYGKVVWNNQBNVHEHQGYTFEADVT 126  
 QY 157 ELVAPGEKRLTIGVNNELTHEFTIPGKITTGNATGRIOITYOHDFPNVAGLARSTWLS 216  
 DB 127 PYVIAQSVARITVCVNNELMWOTIPPGMVT--DENGKKOSYFHDVFNAGHRSVWLYT 185  
 QY 217 VPOOHIDIDIVWVDV--DCDNGILINYEVEVANOQTGQIQISVIDEDGAIYAKASGQSTV 274  
 DB 186 TPTWYDDITVTHVADDCNHAQVDMQV--VAN--GDVSELADADQOVVATGGTSGTL 241  
 QY 275 TIPSVKLMQFGAAYLYQLQVNIYSSGDVVDVTYNLATGVRTVYVAGSQPLNGKPFYFTG 334  
 DB 242 QVNVPHLMQEGEGYLYELCV--TAKSQTECDIYPLRVGIRSVAVKGEQFLINHKPFYFTG 299  
 QY 335 FKGHEDTAVRGKHDPAAYVHDPLQMKWIGANSFRTSHYVYAEVWDFARNQIVVDET 394  
 DB 300 FGRHEDADLRGKGFVDVVLVHDALMDWIGANSYRSHYVYAEVWDFARNQIVVDET 359  
 QY 395 PAVGNLIAL-MGVSESGAP--QTFPPAINDKTQEAHKOAIREFLIARDKNHASYVMMSIAN 452  
 DB 360 AAVGFNLISLGIGFEAGNKKPKELYSEAVNGETQAHQAIKELIARDKNHPSVMSIAN 419  
 QY 453 EPASHEDGAREYEPITNLTRQDPTPTPTFANVGATAYQLDRIIDLFDVSCINRYFGWY 512  
 DB 420 EPDTRPGAREYFAPLAELATRKLDPTPTPTCVNMFCDNHTDITSDLPVLCINRYGYWY 479  
 QY 513 SQTGDLFEAPAALEKELHGWQEKFRPIVWTEYTGADTLAHLISILGPKSEEPQVOMLDM 572  
 DB 480 VQSGDLETAKEVLEKELHGWQEKHPIITTEYGVDTLAGLSMYTDMWSEEQCAMLDM 539  
 QY 573 YHRVPRIRISMGHEHWNFADPOTNIGIIRVDGNKKGVFTDRDKPKYAAHSLRARWTSID 632  
 DB 540 YHRVPRIRISMGHEHWNFADPOTNIGIIRVDGNKKGVFTDRDKPKYAAHSLRARWTSID 599

RESULT 4  
 AAY68840  
 ID AAY68840 standard; protein; 1010 AA.  
 XX

AC AAY68840;

DT 16-MAY-2000 (first entry)

DE Fusion protein of the DNA-binding domain of NAP and GUS.

XX ZAP2; poly(ADP-ribose) polymerase; PARP; poly(ADP-ribose) transferase;  
 KM programmed cell death; apoptosis; growth rate; stress; cold; pathogen;  
 KM pest; drought; heat; fungi; nematode; seed-shatter.

XX Synthetic.  
 OS Arabidopsis thaliana.

PN WO200004173-A1.

PD 27-JAN-2000.

PF 12-JUL-1999; 99WO-EP004940.

PR 17-JUL-1998; 98US-00118276.

PA (PLBZ ) PLANT GENETIC SYSTEMS NV.

PI Babychuk E, Kushnir S, De Block M;

DR WPI; 2000-182436/16.

PT Modulating cell death, growth and stress resistance in eukaryotes.

PT Specifically plants, used, e.g to impart fungus or nematode resistance.

PS Disclosure; Page 112-116; 126pp; English.

XX The present sequence represents a fusion protein of the DNA-binding

CC domain of the NAP protein of Arabidopsis and the GUS protein. The NAP  
 CC protein is a poly(ADP-ribose) polymerase (PARP) protein (also known as  
 CC poly(ADP-ribose) transferase). The PARP enzyme is involved in programmed  
 CC cell death or apoptosis, and is a nuclear enzyme. The NAP polynucleotide  
 CC sequences can be used for modulation of programmed cell death in  
 CC eukaryotic cells. The method is used, specifically in plants, to induce,  
 CC or protect against, programmed cell death, depending on the extent to  
 CC which PARP activity is reduced. Reducing expression of endogenous NAP  
 CC class PARP only is also used to modulate programmed cell death, to  
 CC increase growth rate and to produce plant cells that are more tolerant of  
 CC stresses (cold, chemical treatments, pathogens, pests, drought, heat, etc.,  
 CC or during transformation). Particular applications are generation of  
 CC plants that are resistant to fungi or nematodes; are male or female  
 CC sterile; or have better seed-shatter properties. The methods are also  
 CC used to improve growth of transformed plant cells (and derived calli or  
 CC complete plants)

XX Sequence 1010 AA;

Query Match 51.1%; Score 1714; DB 3; Length 1010;

Best Local Similarity 55.2%; Pred. No. 2, 2e-130;

Matches 331; Conservative 92; Mismatches 161; Indels 16; Gaps 9;

QY 40 VNPRTSSRELVLNDGLMKFAL--ASGLNDTAQPTAPLPKGLCEPPVPSYNDIFISRE 96  
 DB 409 VNPVEPTREIKLDDLMFSLDRENGCIDQ--RMWESALQESRAIAYGSEFNDQPADAD 466  
 QY 97 IDHVGWVYVYOREVYIPKMSOERYLVRAESATHGRIVYNNRLVAEHVGGYTPFEADYT 156  
 DB 467 IRNYAGNVVYQREVFIPKMGAGRIYLRPAVTHYKVVWVNNQVEHGGYTPFEADYT 526  
 QY 157 ELVAPGEKRLITGVNNELTHETIPPGKITTGATGKRIQTYQHDPYNYAGLARSIMLYS 216  
 DB 527 PYVIAKGSYRITVCVNNELMWQITPPGAVIT--DENKKKQSYFHFENYAGIHRSMVLYT 585  
 QY 217 VPOQHIDITVTVDV--DGDNGLINEVEVYANOTGOIQISVIDEGALVAKASGAGQYV 274  
 DB 586 TPWTWDDITVTVTHVAQDCNHSVDMQV--VAN--GDVSELRDADQOVVATGQGTSGTL 641  
 QY 275 TIPSVLMOPGAAUYLOLVNIVGSSGDVDTYNLATGRTVYVAGSOPFLNGKPEYFTG 334  
 DB 642 QVNPRLMWQEGSYLYELCV--TAKSQTECDIYPLRVGRSAVAVGEGFLNHHKPEYFTG 699  
 QY 335 FGHEDTAVRGKGDPAVYVHDFOLMKWIGANSFRTSHYPYAEVWDFADRNIGIYVIDET 394  
 DB 700 FGHEDADLRKGFNDVVLVHDLMDWIGANSYRISHYPYAEEMLDWADDEHGIYVIDET 759  
 QY 395 PAVGLNIAL-MGVSESGAP-QITTPDAINDKTOEAKOAIRELIAADKXHASVVMMSIAN 452  
 DB 760 AAVGFNLISIGIFPEAGNKPELYSEBAVNGETQOAHQAIKELIARDKXHPSVVMSIAN 819  
 QY 453 EPASHEDGAREYEPETLNLTRQLDPTRPITTFANVGATYQLDRISLDFVSCINRYFGWY 512  
 DB 820 EPDTRQGARREYFAPLAETRKLDPTRPITTCVNVVFCDAHTDTISLDFVLCLNRYFGWY 879  
 QY 513 SQTGDLBEAALAEKELHGWOKFRRPIVMTYEGADTLAGLSILGLPMSSEFOVMMLDM 572  
 DB 880 VQSGDLETAKEVLEKELLMQEKLHQPIITTEGVTTLAGLSHMYTDMMSSEVQCAMLDM 939  
 QY 573 YHRVFPRISSMAGEHYVNPADPOTNIGIRVDGNKKGVFTDRKPPAAHSLPARFTSID 632  
 DB 940 YHRVFPRISSMAGEHYVNPADPOTNIGIRVDGNKKGVFTDRKPPAAHSLPARFTSID 632  
 QY 940 YHRVFPRISSMAGEHYVNPADPOTNIGIRVDGNKKGVFTDRKPPAAHSLPARFTSID 632  
 DB 940 YHRVFPRISSMAGEHYVNPADPOTNIGIRVDGNKKGVFTDRKPPAAHSLPARFTSID 632

RESULT 5  
 AAW93827  
 ID AAW93827 standard; protein; 603 AA.

AC AAW93827;

DT 25-JUN-1999 (first entry)

DE E. coli GUS protein fragment.

XX GUS; BcGUS; beta-glucuronidase; secreted; reporter molecule; marker;  
 KW receptor molecule; diagnostic tool; transgene construction; plant;  
 KW insect; cleavage; detoxification; glucuronide.

XX Escherichia coli.

XX WO9913085-A2.

XX 18-MAR-1999.

XX 09-SEP-1998; 98WO-US019217.

XX 09-SEP-1997; 97US-0058263P.

XX (CMB-) CAMBIA BIOSYSTEMS LLC.

XX Jefferson RA, Kilian A, Keese PK;

XX WPI; 1999-229241/19.

XX New isolated microbial beta-glucuronidase.

XX Example 3; Fig 13D; 76pp; English.

CC This invention describes a novel secreted form of Bacillus sp. beta-  
 CC glucuronidase (BcGUS). The microbial BcGUS polypeptide can be used as a  
 CC reporter/effector molecule and as a diagnostic tool. The products of the  
 CC invention can be used as markers for transgene constructions, e.g. in  
 CC plants or insects. They can also be used for the cleavage and  
 CC detoxification of glucuronides and to examine conjugation patterns of  
 CC glucuronides

XX Sequence 603 AA;

Query Match 51.0%; Score 1711; DB 2; Length 603;

Best Local Similarity 55.0%; Pred. No. 1, 8e-130;

Matches 330; Conservative 93; Mismatches 161; Indels 16; Gaps 9;

QY 40 VNPRTSSRELVLNDGLMKFAL--ASGLNDTAQPTAPLPKGLCEPPVPSYNDIFISRE 96  
 DB 2 LRVETPTREIKLDDLMFSLDRENGCIDQ--RMWESALQESRAIAYGSEFNDQPADAD 59  
 QY 97 IDHVGWVYVYOREVYIPKMSOERYLVRAESATHGRIVYNNRLVAEHVGGYTPFEADYT 156  
 DB 60 IRNYAGNVVYQREVFIPKMGAGRIYLRPAVTHYKVVWVNNQVEHGGYTPFEADYT 119  
 QY 157 ELVAPGEKRLITGVNNELTHETIPPGKITTGATGKRIQTYQHDPYNYAGLARSIMLYS 216  
 DB 120 PYVIAKGSYRITVCVNNELMWQITPPGAVIT--DENKKKQSYFHFENYAGIHRSMVLYT 178  
 QY 217 VPOQHIDITVTVDV--DGDNGLINEVEVYANOTGOIQISVIDEGALVAKASGAGQYV 274  
 DB 179 TPWTWDDITVTVTHVAQDCNHSVDMQV--VAN--GDVSELRDADQOVVATGQGTSGTL 234  
 QY 275 TIPSVLMOPGAAUYLOLVNIVGSSGDVDTYNLATGRTVYVAGSOPFLNGKPEYFTG 334  
 DB 235 QVNPRLMWQEGSYLYELCV--TAKSQTECDIYPLRVGRSAVAVGEGFLNHHKPEYFTG 292  
 QY 335 FGHEDTAVRGKGDPAVYVHDFOLMKWIGANSFRTSHYPYAEVWDFADRNIGIYVIDET 394  
 DB 293 FGHEDADLRKGFNDVVLVHDLMDWIGANSYRISHYPYAEEMLDWADDEHGIYVIDET 352  
 QY 395 PAVGLNIAL-MGVSESGAP-QITTPDAINDKTOEAKOAIRELIAADKXHASVVMMSIAN 452  
 DB 353 AAVGFNLISIGIFPEAGNKPELYSEBAVNGETQOAHQAIKELIARDKXHPSVVMSIAN 412  
 QY 453 EPASHEDGAREYEPETLNLTRQLDPTRPITTFANVGATYQLDRISLDFVSCINRYFGWY 512  
 DB 413 EPDTRQGARREYFAPLAETRKLDPTRPITTCVNVVFCDAHTDTISLDFVLCLNRYFGWY 472  
 QY 513 SQTGDLBEAALAEKELHGWOKFRRPIVMTYEGADTLAGLSILGLPMSSEFOVMMLDM 572  
 DB 513 SQTGDLBEAALAEKELHGWOKFRRPIVMTYEGADTLAGLSILGLPMSSEFOVMMLDM 572

Db	473	VOSGDLFETAEKVLKELKELLAMQEKHLQPIITREYGVDTLGLHSMVTDMMSEETYGCAMLDM	532
Qy	573	YHRVFPDRRESNAGEYVWNPADPQNTLGIITRDGNKKGVFTDRDKKAAAHASIRAAWTSID	632
Db	533	YHRVFEDRVSAAVGEQVWNPADFATISQGIILRVGNNKKGIPTDRKPKSAAFLLQKSWTGMN	592
RESULT 6			
ID	AAW93824	standard; protein; 603 AA.	
AC	AAW93824;		
XX			
DT	25-JUN-1999	(first entry)	
XX			
DE		Human GUS protein.	
XX			
KW	GUS; BCGUS; beta-glucuronidase; secreted; reporter molecule; marker;		
KM	receptor molecule; diagnostic tool; transgene construction; plant;		
KM	insect; cleavage; detoxification; glucuronide; human.		
XX			
OS	Homo sapiens.		
XX			
PN	W09913085-A2.		
XX			
XX	18-MAR-1999.		
PF	09-SEP-1998;	98WO-US019217.	
XX			
PR	09-SEP-1997;	97US-0058263P.	
XX			
PA	(CAMB-) CAMBIA BIOSYSTEMS LLC.		
XX			
PI	Jefferson RA, Kilian A, Keese PK,		
DR	WPI; 1999-229241/19.		
XX			
XX	New isolated microbial beta-glucuronidase.		
XX			
PS	Example 1; Fig 5; 76pp; English.		
XX			
CC	This invention describes a novel secreted form of Bacillus sp. beta-		
CC	glucuronidase (BogUS). The microbial BogUS polypeptide can be used as a		
CC	reporter/effector molecule and as a diagnostic tool. The products of the		
CC	invention can be used as markers for transgene constructions, e.g. in		
CC	plants or insects. They can also be used for the cleavage and		
CC	detoxification of glucuronides and to examine conjugation patterns of		
CC	glucuronides		
SQ	Sequence 603 AA;		
Qy	Query Match	51.0%; Score 1711; DB 2; Length 603;	
Db	Best Local Similarity	55.0%; Pred. No. 1,8e-130;	
Matches	330; Conservative	93; Mismatches 161; Indels 16; Gaps 9	
Qy	40 VRPORTSSRELVNLDGLMKFAL--ASGLNDTAPQWTPALPKGLGECPPASYNIDIFISRE	96	
Db	2 LRPEVETPREIKKLDGLWAFSLDRNCGIDQ--RWESALQESRAIAVPGSFNDQFADAD	59	
Qy	97 IHDHVGWYTYQREVLPKGMSEERILVAESATHHGRITYANNRLVAEHVGGTTPREADV	156	
Db	60 IRNVAAGNVYOREVAPIPKGMAGQRIVLRFDAVTHGKAVNNQEVHEHGGTTPREADV	119	
Qy	157 ELVAPGEKFRLLTIGVNNELTHEITPPGKITTGNAIGKRIQTYQHDFFVNVAGLARISIMLYS	216	
Db	120 PYVLAGSKSVRIITVCVNNELNMOTITPPGVIT--DENGKKKQSTFFHDFPNYAGIHRSVMLYT	178	
Qy	217 VPOOHIDITVTDV--DENGGLINVEYEVANQTTGOIQISVYDEGAIIVAKASGAOGTV	274	
Db	179 TPNTWVDVITVTVTHVADQCNHASVDMQV--VAN--GDVSVELRADQGVAVANGQGTSGTL	234	
Qy	275 TTPSVKLMQPGAAIYIQLVNIIVGSSGDVDTYNIATGVRTYTKVAGSGLNGKFPFTTG	334	

Dd	235	QVNVPHLMQPEGYLVELCV--FAKSQTECDIPLRFGVIGSVAVKGEQFLINHKPFYFTG	292
Qy	335	FGKKEEDVAVKGGKDPAYMHDPOLMKYIGANSRTSHYTPYAEVMPADBNGLVYIDET	394
Dd	293	FGREHEDDLRGKGFDPNVLWHDALMDWIGANSYRSHHPYAEEMLDMADEHGIVYIDET	352
Qy	395	PAVELNLIAL-MGVSESGAP-OTFPDPAINDKTOEHAHQALRELARDKNHASYVMGSIAN	452
Dd	353	AAVFENSLSLGIGFAGNKKPKVELSEEVNGETOQALQALKELIARDKNHPSVVMGSIAN	412
Qy	453	EPASHEDEGAREYEPPLTNLTROLDPTRPITFANVGATYOLDRIISDLFDVSCINRYGMY	512
Dd	413	EPDTRPGAGREYFAPPLAEATRKLDPTRPICVANNMFCDAHTDITSDLFVLCINRYGMY	472
Qy	513	SGTCDDLEBAALAEKELHGMQEKHFRPIVTETGADTLAGLHSLGLPNSGEFQVOMLDM	572
Dd	473	VQSGDLELTAEKVELKEKELLAWOEKTHQPIITTEYVDTLAGLSHYTDMMSSEBYCAWMLDM	532
Qy	573	YHRFEDLESAGSHVWNPADPOTNIGITIVDNGKKGVFTRDRKPKAASLTARMSID	632
Dd	533	YHRFVDEVSAAVGEQVNNFPADPATSQGLIRVGKKKGI FTRDRKPKASALFLQKRMGMN	592
RESULT 7			
ID	AAB28431	AAB28431 standard; protein: 603 AA.	
XX	AC	AAB28431;	
XX	DT	26-JAN-2001 (first entry)	
DE	XX	Human beta-glucuronidase HGUS.	
XX	KW	Microbial; beta-glucuronidase; GUS; Enterobacter; Salmonella;	
KV	KW	Pseudomonas; Staphylococcus; Thermocoga; transgenic plant; bioindicator;	
XX	XX	transgenic insect; marker; glucuronide detoxification.	
OS	XX	Homo sapiens.	
FN	XX	W0200055333-A1.	
PD	XX	21-SEP-2000.	
PF	XX	16-MAR-2000; 2000W0-US007107.	
PR	XX	17-MAR-1999; 99US-00270957.	
PA	XX	(CAMB-) CAMBIA BIOSYSTEMS LLC.	
P1	XX	Jefferson RA, Mayer JE;	
P1	XX	WPI; 2000-647075/62.	
PT	XX	Novel microbial beta-glucuronidase genes and gene products used as	
PT	XX	targeter/effecter molecule, as diagnostic tool, in positive selection, to	
PT	XX	targeter molecules to specific cells and to detect and track linked genes.	
PS	XX	Example 4; Fig 5A; 116pp; English.	
CC	XX	The present sequence is a microbial beta-glucuronidase (GUS) protein. GUS	
CC	XX	genes were obtained from six different genera: Enterobacter/Salmonella,	
CC	XX	Pseudomonas, Staphylococcus and Thermocoga. Microbial GUS can	
CC	XX	be used as a reporter/effecter molecule for transgenic constructions and	
CC	XX	in vitro diagnostic applications. It may also be used to generate	
CC	XX	sentinel plants that serve as bioindicators of environmental status. It	
CC	XX	may be used to generate transgenic insects for tracking insect	
CC	XX	populations or to facilitate the development of a bioassay for compounds	
CC	XX	that affect molecules critical for insect development (e.g. juvenile	
CC	XX	hormone). Secreted GUS may also serve as a marker for beneficial fungi	
CC	XX	destined for release into the environment. In animal systems, secreted	
CC	XX	GUS may be used to achieve extracellular detoxification of glucuronides	
CC	XX	(e.g. toxin glucuronide) and to examine conjugation patterns of	
CC	XX	glucuronides. Microbial GUS may also be used in traditional medical	





[illegible]

QY	217	VPGQHIQDITVTDV--DGDNLLIVEYEVAQTGGIOISVIDEGALVVAASAGQGV	214
Db	179	TPTNVADDTITVTHVAQDCNHAUVQV-VAN---GDVSELRADQOVAVGQTSGL	234
QY	275	TIPSVKMAOPGAAYLYLOLVNIVGSSGDVDTYNLATGVRTYKVASGQPLNGKPEYFTG	334
Db	225	QVNVNPHLMQPGEGYLVELCV--TAKSQTECDIYPLRVGIRSAVAKGEQFLNNKPFYFTG	292
QY	335	FGKHEDTVVRGKGHPAYMVDFOLMKMI GANSFRTSHFPYABEVDVFDNRGIVVIDE	394
Db	293	FGRHEDADLRKGEPNVNLMVHDHALMDVIGANSYRTSHFPYABEMLDMADDEHIVVIDE	352
QY	395	PAYGINLIL--NGVSSSGAP-QTFETDAILNDKIQEAKHKAIRELILARDKNHASVVMSSIAN	452
Db	353	AAGVGNLSLGI GFPAKGNPKELYSSEAVNGEYQAHLQIKELIARDKNHPSVVMSSIAN	412
QY	453	EPASHEDAREYFEPLTNLTROLDPTRPITFANVGATATQDLRISDLFPVSCINRYFGWY	512
Db	413	EPDTRPGQAREYFAPLAETRLDPTTRPTTCVNWFCDAHTDITISLFPVLCLNRYGWY	472
QY	513	SQTGDLFEAALAEKELHGMOKFHPVMTVEYGADTLAGLHSILGLPWSSEFQOVMLD	572
Db	473	VOSGDLFEAEKVELKEKELLAMOKLHQPIITTEYGVDTLAGLHSMYDMSSEYQAAMD	532
QY	573	YHRVDRDRIESMAGEVWNNPAPQOTNLGIIIRVGNKKGVFTTRPKPKAAHSLPARKTSID	632
Db	533	YHRVDFRVSAAVGEQVWNNPADFATISQGLIRVGNKKGIPTTRPKPKSAAFLQKRWGMN	592
RESULT 10			
ABJ19649	ID	ABJ19649 standard; protein; 603 AA.	
XX	AC	ABJ19649;	
XX	DT	03-APR-2003 (first entry)	
DE	XX	Artificial plant chromosome related beta-glucuronidase SEQ ID No 17.	
XX	XX	Plant artificial chromosome; PAC; transgenic plant; vaccine;	
KM	XX	blood factor; herbicide; stress; agronomical; nutrient quality;	
KW	XX	bacterial artificial chromosome; BAC; yeast artificial chromosome; YAC;	
XX	XX	enzyme.	
XX	OS	Escherichia coli.	
XX	PN	MO200296923-A1.	
PD	XX	05-DEC-2002.	
PF	XX	30-MAY-2002; 2002WO-US017451.	
PR	XX	30-MAY-2001; 2001US-0294687P.	
PR	XX	04-JUN-2001; 2001US-0296329P.	
PA	XX	(CHRO-) CHROMOS MOLECULAR SYSTEMS INC.	
PA	XX	(AGRI-) AGRISOMA INC.	
PI	XX	Perez C, Fabijanski SF, Perkins E;	
DR	XX	WPI; 2003-140436/13.	
DR	XX	N-PSDB; ABT16606.	
PT	XX	Producing artificial chromosome by introducing a nucleic acid into plant	
PT	XX	cell, selecting artificial chromosome that has one or more repeat regions	
PT	XX	with equivalent amounts of euchromatic and heterochromatic nucleic acids.	
PS	XX	Disclosure; Page 245-246; 269pp; English.	
CC	XX	The invention relates to a novel method for producing plant artificial	
CC	XX	chromosomes. The invention also relates to methods for targeting	
CC	XX	insertion of heterologous DNA into plant artificial chromosomes, methods	
CC	XX	for delivery of plant chromosomes to selected cells and tissues. The	



xx	The invention relates to a method of expressing a heterologous
cc	polypeptide by a host cell. The methods are useful for producing altered
cc	seed meals by manipulating oil bodies of plants. The present sequence is
cc	used in the exemplification of the present invention.
xx	
SQ	Sequence 604 AA:
Query Match	50.9%; Score 1706; DB 7; Length 604;
Best Local Similarity	54.8%; Pred. No. 4.6e-110;
Matches 329; Conservative 94; Mismatches 161; Indels 16; Gaps 9;	
Oy	40 VRPORTSSRELVMNDLGMKFAL--ASGLDTPAQPTAPLPKGECPVSPASYNDFISRE 96   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   : 3 LRPETRTRELKKLDGDMAFSLDENCGIQ-RWMSAQAQSALINVPSPFNQPADAD 60
Oy	97 IHDHGVVVYTOREVIVPKGSQERYLVRAESATTHGRITYYNNRLVAHVGGYTFFEDVT 156   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   : 61 IRNVAGVWVOOREVFIEFKWAGORIVRFPFVATHYGKVWNNOEWMHGOGYTFPEDAVT 120
Oy	157 ELVAPGEKFRLLTVNNNELTHETRIPPKKITGTGNATGRIOTYGHDPFNVGARSIMYS 216   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   : 121 PYTAGSASRVITTCVNNELEMKQTILPPGWAIT-DENGKKKGSYFHDFPNYAGIHRSWLYT 179
Oy	217 VPQHODIDITVDYV--DGNGGLINTVEVANQTTGOIGISVIDEGAIYAKASGAOQTV 274   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   : Db TPNTMTWDIDITVTHVADQCNHASVDQM-VAN--GPVSVELRADDOOVAVATQGTSCTL 235
Oy	275 TIESVKLMOPKAAILYLQLOVINIVSSSGDVVDYTNLATAGVRTVKVASGFPLNKPEFYFTG 334   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   : Db QVNPFLHMOPPEGLYELCV-TAKSTGETCIDLPLRVGISNSAAVKGQOFLLNRKPFTFG 293

Db	294	FGRIEDADLCKGKGDVNTLAWHDHALMDMIGANSYRSHYPFAEBEMLDMADEHGIIVTDET	353
Oy	395	PAVLGNIAL-NCVSESGAP-QTFTPDAINDKTQEAHKQAIRBELIARDKNTASVVMMSIAN	452
Db	354	AAVGFSLSLGIFGPAKNKPKEYSEEAVNGETQAHQAIKELIARDKNHPSVVMMSIAN	413
Oy	453	EPASHEDGAREYFEPPLTNLTROLDPTPTPIPTPANYGTATYQLDRLSDLFDVSCINRYRGWY	512
Db	414	EPDTRPGAREYFAPLAEATRKLDPTPTPTICVNMPCDAHTDTISLFDVLCINRYRGWY	473
Oy	513	SQTDDLEAEALAEKELHGWQEKHPRPIVMTVEYGADTLAGHSITGLPMSEEPQOVMLDM	572
Db	474	VQSGDLEFAEKVLEKELLAMQEKLHOPILITTEYGVDTLAGHSWYTDPMSEEQCAMLDM	533
Oy	573	YHRVFDRIESWAGEVWNNFADPOTNLGIIRVDGKKKKGVFTDRKPKKAAASLSLRARWTSID	632
Db	534	YHRVFDKVSAAVGEQVWNNFADPATSQGLIRVGKKKGIPTDRKPKKSAFLQGRWTCMN	593
RESULT 13			
ADD27989			
ID	ADD27989	standard; protein; 659 AA.	
XX	ADD27989;		
XX	AC		
XX	DT	15-JAN-2004 (first entry)	
XX	DE	Oleosin/beta-glucuronidase fusion protein #2.	
XX	KW	beta-glucuronidase; oleosin; thale cress; altered seed meal; oil body;	
XX	XX	plant; enzyme.	
OS	OS	Escherichia coli.	
XX	XX	Arabidopsis thaliana.	
PN	US2003126631-A1.		
XX	PD	03-JUL-2003.	
XX	XX		



```

Db      482 QVNPPLMOPGEGYLVELCV--TAKSQTECDIYPLAVGIRSVAVKGOQFLNKKPFYFTG 539
Qy      335 FKGHEBTAVRGKGDPAVYVHDFOLMKWIGANSFRTSHYPYAEVNDPADRNQIVVIDET 394
Db      540 FGRHEDADLRGKGFDPVNLVHDLMDWIGANSYRTSHYPYAEMLMADDEHGIIVIDET 599
Qy      395 PAVGLNIAL-MGVSESGAP-QFTTPDAINDKTOEAKHKAIRLIARDKXHASVVMMSIAN 452
Db      600 AAVGFNLISLIGFEAGNKKPELYSEEAUVNGETOQAHQAIKELIARDKXHASVVMMSIAN 659
Qy      453 EPASHEDGAREYFEPLTNLTROLDPTRPITFANVGTATYQOLDRIISDFVSCINRYFGWY 512
Db      660 EBDTRQGARREYFAPLAETARKDPTRPITCVVMFCDAHTDTISDLFDVLCLNRYFGWY 719
Qy      513 SQTGLEEBAALKEKELHGMOKFHRPIVWTEYGADTLAGLSILGLPMSSEPOVOMLDM 572
Db      720 VOSGDLETAKEKLEKELHGMOKFHRPIIITEYGVDTLGLHSMYDMMSSEPOCAMLDM 779
Qy      573 YHRVPDRISMAGEHYWNPADPOTNLGIIRVDGNKKGVFTRDKKPKAAHSLAARTSID 632
Db      780 YHRVPDRISAVVGEQVWNPADPOTNLGIIRVDGNKKGVFTRDKKPKAAHSLAARTSID 839

RESULT 15
ABBB4107
ID      ABB84107 standard; protein; 603 AA.
XX      ABB84107;
AC      17-SEP-2002 (first entry)
DT      GUS protein #1.
XX      GUS; refractory; beta-glucuronidase; screening; ds.
XX      Unidentified.
XX      CN1338515-A.
XX      06-MAR-2002.
XX      18-AUG-2000; 2000CN-00119633.
XX      18-AUG-2000; 2000CN-00119633.
XX      (REBI-) RES CENT BIO TECHNOLOGY SHANGHAI ACAD AG.
XX      Yao Q, Xiong A, Peng R;
XX      WPI; 2002-384278/42.
XX      N-PSDB; ABL61358.
XX      Refractory beta-glucosiduronatase gene and its obtaining process.
XX      Claim 1; Page 1-3 (Claims); 23pp; Chinese.
XX      This invention describes a novel refractory beta-glucuronidase (GUS) gene
XX      prepared through DNA mutation. The gene is used in the creation of an
XX      expression carrier, which is transferred into a colibacillus. The GUS
XX      gene has refractory power (85 degrees C) and can be used to screen
XX      transgenic plants with high efficiency. This sequence represents a GUS
XX      protein described in the disclosure of the invention
XX      SQ      Sequence 603 AA;

```

Query Match 50.6%; Score 1697; DB 5; Length 603;

Best Local Similarity 54.5%; Pred. No. 2.5e-129; Matches 327; Conservative 95; Mismatches 162; Indels 16; Gaps 9;

```

Qy      40 VRRPQTSSEELVNLGLMKFAL--ASGLNDTAQPTAPLPKGLBCPPVASYNDIFISRE 96
Db      2 LRPVETPTREIKLGLDGLMAFSLDRNCGIDQ--RWMESALQESRAIAPGSPFNDQFADAD 59

```

```

Qy      97 IDHVGWVYVYOREBIVPKMSOERYLVRAESATTHGRIVYNNRLVAEHGYTTPFEADYT 156
Db      60 IRRYAGNVYQOEVEPLPKMAGORIVLRPDATYHKKVWVNNQEVNHEGQITTFPADYT 119
Qy      157 ELVAPGEKRLTIGVNNELTHETIPPKITTGNGATKRIQYQHDYFNAGLARSIMLYS 216
Db      120 PVIACKSVRITVCVNNELMOTIPFGWIT--DENCKQKSYHDFPNVAGIHRSVLWLT 178
Qy      217 VPGQHIQDITTVTDV--DGDNGILNVEEVANOTQIQISVYDEGALVAKASGAGQY 274
Db      179 TPWTWDDITVTVTHVAQDCNHASVDMQV--VAN--GDVSEVLADADQOVVATGQGTSGTL 234
Qy      275 TIPSVYLMOPGAAYLYOLOVNIYSSGVDVDFYNLATGRTYVAVASOPFLNKKPFYFTG 334
Db      235 QVNPPLMOPGEGYLVELCV--TAKSQTECDIYPLAVGIRSVAVKGOQFLNKKPFYFTG 292
Qy      335 FKGHEBTAVRGKGDPAVYVHDFOLMKWIGANSFRTSHYPYAEVNDPADRNQIVVIDET 394
Db      293 FGRHEDADLRGKGFDPVNLVHDLMDWIGANSYRTSHYPYAEMLMADDEHGIIVIDET 352
Qy      395 PAVGLNIAL-MGVSESGAP-QFTTPDAINDKTOEAKHKAIRLIARDKXHASVVMMSIAN 452
Db      353 AAVGFNLISLIGFEAGNKKPELYSEEAUVNGETOQAHQAIKELIARDKXHASVVMMSIAN 412
Qy      453 EPASHEDGAREYFEPLTNLTROLDPTRPITFANVGTATYQOLDRIISDFVSCINRYFGWY 512
Db      413 EBDTRQGARREYFAPLAETARKDPTRPITCVVMFCDAHTDTISDLFDVLCLNRYFGWY 472
Qy      513 SQTGLEEBAALKEKELHGMOKFHRPIVWTEYGADTLAGLSILGLPMSSEPOVOMLDM 572
Db      473 VOSGDLETAKEKLEKELHGMOKFHRPIIITEYGVDTLGLHSMYDMMSSEPOCAMLDM 532
Qy      573 YHRVPDRISMAGEHYWNPADPOTNLGIIRVDGNKKGVFTRDKKPKAAHSLAARTSID 632
Db      533 YHRVPDRISAVVGEQVWNPADPOTNLGIIRVDGNKKGVFTRDKKPKAAHSLAARTSID 592

```

Search completed: March 18, 2005, 23:48:06  
Job time : 174 secs

**THIS PAGE BLANK (UPR)**